

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12Q 1/68, 1/70, C12P 19/34, C07H 21/04, 21/02, A61K 41/00</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/50582</b> <b>(43) International Publication Date:</b> 12 November 1998 (12.11.98)
<b>(21) International Application Number:</b> PCT/US97/12722 <b>(22) International Filing Date:</b> 21 July 1997 (21.07.97) <b>(30) Priority Data:</b> 08/853,522                      8 May 1997 (08.05.97)                      US <b>(71) Applicant:</b> CALIFORNIA INSTITUTE OF TECHNOLOGY [US/US]: 1201 E. California Boulevard, Pasadena, CA 91125 (US). <b>(72) Inventor:</b> DERVAN, Peter; 1235 St. Albans Road, San Marino, CA 91108 (US). <b>(74) Agents:</b> BERNTSEN, Denis, A. et al.; McDonnell Boehnen Hulbert & Berghoff, 300 South Wacker Drive, Chicago, IL 60606 (US).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i>
<b>(54) Title:</b> COMPLEX FORMATION BETWEEN dsDNA AND OLIGOMER OF HETEROCYCLES  <b>(57) Abstract</b>  Methods and compositions are provided for forming complexes between dsDNA and oligomers of heterocycles, aliphatic amino acids, particularly omega-amino acids, and a polar end group. By appropriate choice of target sequences and composition of the oligomers, complexes are obtained with low dissociation constants. The formation of complexes can be used for identification of specific dsDNA sequences, for inhibiting gene transcription, and as a therapeutic for inhibiting proliferation of undesired cells or expression of undesired genes.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**Complex Formation Between dsDNA  
and Oligomer of Heterocycles**

The U.S. Government has certain rights in this invention pursuant to Grant Nos.  
5 GM 26453, 27681 and 47530 awarded by the National Institute of Health.

**CROSS REFERENCE TO RELATED APPLICATIONS**

This application is a continuation-in-part of application serial no. 08/837,524, filed  
21, April 1997, which is a continuation-in-part of application serial no. 08/607,078, filed  
10 February 26, 1996, filed as PCT application US97/03332, on February 20, 1997, and  
provisional application serial nos. 60/023,309, filed on July 31, 1996, 60/024,374, filed  
on August 1, 1996, 60/026,713 filed on September 25, 1996, and 60/038,384, filed on  
February 14, 1997.

15 **INTRODUCTION**

**Background**

With the explosion of techniques for the synthesis, analysis and manipulation of  
nucleic acids, numerous new opportunities have arisen in diagnostics and therapeutics. In  
research there is substantial interest in being able to identify DNA sequences, which may  
20 be associated with specific organisms, alleles, mutations, and the like, to understand  
particular genetic processes, to identify diseases, for forensic medicine, etc. Also, for  
many purposes, one may wish to modulate the activity of a particular gene, so as to  
identify the function of a particular gene, the effect of changes in its cellular concentration  
of its gene product on the function of the cell, or other cellular characteristic. In  
25 therapeutics, one may wish to inhibit the proliferation of cells, such as bacterial, fungal  
and chlamydia cells, which may act as pathogens, of viruses, of mammalian cells, where  
proliferation results in adverse effects on the host, or other situation. In vivo, one may  
provide for reversible or irreversible knock out, so that information can be developed on

the development of a fetus, or the effect on the organism of reduced levels of one or more genetic products.

In a number of seminal papers, Peter Dervan's group has shown that oligomers of  
5 nitrogen heterocycles can be used to bind to dsDNA. It has been shown that there is  
specificity in that G/C is complemented by N-methyl imidazole (Im)/ N-methyl pyrrole  
(Py), C/G is complemented by Py/Im, A/T and T/A are redundantly complemented by  
Py/Py. In effect, N-methyl imidazole tends to be associated with guanosine, while N-  
methyl pyrrole is associated with cytosine, adenine, and thymidine. By providing for two  
10 chains of the heterocycles, as 1 or 2 molecules, a 2:1 complex with dsDNA is formed,  
with the two chains of the oligomer antiparallel, where G/C pairs have Im/Py in  
juxtaposition, C/G pairs have Py/Im, and T/A pairs have Py/Py in juxtaposition. The  
heterocycle oligomers are joined by amide carbamyl groups, where the NH may  
participate in hydrogen bonding with nitrogen and oxygen unpaired electrons of adenine  
15 and thymidine in the minor groove (Figure 1), particularly of adenine. While the  
complexes were of substantial interest, the binding affinities for the most part were less  
than about  $10^6 M^{-1}$ . Furthermore, the discrimination between a target DNA sequence, and  
one involving a mismatch was frequently not better than about two-fold. Therefore, for  
many purposes, the complexes had limited utility.

20  
Improvements in affinity were shown for a cyclic dimer, where the two oligomers  
were joined at their ends by  $\gamma$ -aminobutyric acid, where the affinity was shown to be  
enhanced to about  $10^9 M^{-1}$ . However, the difference in affinity between the target sequence  
and were less than three-fold difference for three different single-base mismatch  
25 sequences. This low sequence-selectivity would severely limit the applications for the  
compound in the presence of a large amount of naturally occurring dsDNA.

Also, for many applications, one wishes to be able to use the sequences with  
viable cells. There was no showing that these oligomers would be capable of being  
30 transported across a cellular membrane to the nucleus and, upon successful transport to the  
nucleus, they could bind to the chromosomal DNA, where the chromosomal DNA is  
present as nucleosomes.

Relevant Literature

- Wade et al., J.AM.CHEM.SOC., 1992, 114, 8783-8794; Mrkish et al.,  
 PROC.NATL.ACAD.SCI. USA, 1992, 89, 7856-7590; Mrkish and Dervan,  
 5 J.AM.CHEM.SOC., 1993, 115, 2572-2576; Wade et al., Biochemistry, 1993, 32, 11385-  
 11389; Mrkish and Dervan, J.AM.CHEM.SOC., 1993, 115, 9892-9899; Dwyer et al.,  
 J.AM.CHEM.SOC., 1993, 115, 9900-9906; Mrkish and Dervan, J.AM.CHEM.SOC.,  
 1994, 116, 3663-3664; Mrkish et al, J.AM.CHEM.SOC., 1994, 116, 7983-7988; Mrkish  
 and Dervan J.AM.CHEM.SOC., 1995, 117, 3325-3332; Cho et al.,  
 10 PROC.NATL.ACAD.SCI. USA, 1995, 92, 10389-10392; Geierstanger, Nature Structural  
 Biology, 1996, 3, 321-324; Parks et al., J.AM.CHEM.SOC., 1996, 118, 6147-6152;  
 Parks et al., J.AM.CHEM.SOC., 1996, 118, 6153-6159; Baird and Dervan,  
 J.AM.CHEM.SOC., 1996, 118, 6141-6146; Swalley et al., J.AM.CHEM.SOC., 1996,  
 118, 8198-8206; Trauger et al., J.AM.CHEM.SOC., 1996, 118, 6160-6166; Szewczyk et  
 15 al., J.AM.CHEM.SOC., 1996, 118, 6778-6779; Trauger et al., Chemistry & Biology,  
 1996, 3, 369-377; Trauger et al., Nature, 1996, 382, 559-561; Kelly et al.,  
 PROC.NATL.ACAD.SCI. USA, 1996, 93, 6981-6985; Szewczyk et al.,  
 ANGEW.CHEM.INT.ED.ENGL., 1996, 35, 1487-1489; Pilch et al.,  
 PROC.NATL.ACAD.SCI. USA, 1996, 93, 8306-8311; White et al., Biochemistry, 1996,  
 20 35, 12532-12537.

SUMMARY OF THE INVENTION

- Methods and compositions are provided for selectively producing a complex at a  
 25 concentration of  $\leq 1\text{nM}$ , between dsDNA and an oligomer of organic cyclic groups,  
 wherein at least 60% of the cyclic groups are heterocyclics, and at least 60% of the  
 heterocycles have at least one nitrogen annular member. The heterocycles form  
 complementary pairs, where at least two of the nucleotide pairs are preferentially paired  
 with a specific pair of heterocycles. There are at least three complementary pairs of  
 30 organic cyclic groups in the complex, either as a result of a hairpin turn in a single  
 oligomer, or the complementation between organic cyclic groups of two oligomeric  
 molecules. Usually, a small aliphatic amino acid will be interspersed in or divide what  
 would otherwise be a chain of six or more consecutive organic cyclic groups. To further  
 enhance binding, a terminus may have at least one aliphatic amino acid of from two to six

carbon atoms and/or an alkyl chain having a polar group proximal to the linkage of the alkyl chain. By appropriate selection of the target sequence, the complementary pairs, unpaired organic cyclic groups, the aliphatic amino acids, and the polar-substituted alkyl chain, complexes may be formed with high affinity, low dissociation constants, and significant disparities in affinity between the target sequence and single-base mismatches. Modifications of the oligomers are used to provide for specific properties and are permitted at sites which do not significantly interfere with the oligomers positioning in the minor groove. The compositions are found to be able to enter viable cells and inhibit transcription of genes comprising the target sequence, cleave at particular sites, become covalently bonded at specific sites, direct selected molecules to a target site, as well as perform other activities of interest.

The oligomers may be combined with dsDNA under complex forming conditions to form the complex. Formation of the complex can be used in diagnosis to detect a specific dsDNA sequence, where the oligomers may be labeled with a detectable label, to reversibly or irreversibly "knock out" genes in vitro or in vivo, cytohistology, to inhibit proliferation of cells, both prokaryotic and eukaryotic, and the like.

#### BRIEF DESCRIPTION OF THE DRAWINGS

20

Figure 1(a) Model for Reconition of the DNA Minor-Groove. The DNA double helix consists of A,T and G,C base pairs like rungs on a twisted ladder. Individual sequences may be distinguished by the pattern of hydrogen bond donors and acceptors displayed on the edges of the base pairs. The A,T base pair presents two symmetrically placed hydrogen bond acceptors in the minor groove, the purine N3 and the pyrimidine O2 atoms represented as circles with dots. The G,C base pair presents these two acceptors, but in addition presents a hydrogen bond donor, the 2-amino group of guanine represented as a circle containing an H. Because the hydrogen bond vector lies towards the guanine-containing strand, GC and CG base pairs may be distinguished in the minor groove. (b) Pairing Rules for DNA Recognition by Pyrrole-Imidazole for side-by-side complexes Py/Impolyamides in the minor groove of DNA, the DNA binding sequence specificity depends on the sequence of side-by-side amino acid pairings. A pairing of Im opposite Py targets a GC base pair while a pairing of Py opposite Im targets a CG base pair. A Py/Py combination is degenerate targeting both AT and TA base pairs. Specificity for

G,C base pairs results from the formation of a putative hydrogen bond between the imidazole N3 and the exocyclic amino group of guanine. Putative hydrogen bonds are represented by dashed lines.

- 5           Figure 2 depicts binding models for (a) 5'-AGTACT-3' in complex with ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp 1 (match) and ImPyPyPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp 2 (mismatch), and (b) 5'-AGTATT-3' in complex with ImPyPyPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp 2 (match) and ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp 1 (mismatch). Circles with dots represents lone pairs on N3 of purines and O2 of pyrimidines, and circles containing H represent the N2 hydrogen of
- 10           guanine. Putative hydrogen bonds are illustrated by dashed lines. The dark and open circles represent imidazole and pyrrole rings, respectively, the curved line represents  $\gamma$ -aminobutyric acid, and the diamond represents  $\beta$ -alanine. Single hydrogen bond mismatches are highlighted.
- 15           Figure 3.(a) (left) Model of nine zinc finger protein TFIIIA with the 5S RNA gene internal control region (ICR). (middle) Sequence of the ICR recognized by zinc finger 4 in the minor groove. (right) Complex of the hairpin polyamide ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp 1 with its target site, 5'-AGTACT-3'. Circles with dots represent lone pairs on N3 of purines and O2 of pyrimidines. Circles containing an H represent the N2 hydrogen of
- 20           guanine. (b) Structures of polyamides ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp (1), ImPyPyPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (2), and ImPyImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (3). (Dp = dimethylaminopropylamide). (c) Binding models filled and open circles represent imidazole and pyrrole rings, respectively, the curved line represents  $\gamma$ -aminobutyric acid ( $\gamma$ ), and the diamond represents  $\beta$ -alanine ( $\beta$ ). Hydrogen bond mismatches are
- 25           highlighted.

### DESCRIPTION OF THE SPECIFIC EMBODIMENTS

- 30           The subject invention provides novel oligomers for forming high affinity complexes with dsDNA. The oligomers comprise organic cyclic groups joined together by short linkers, which oligomers fit in the minor groove of dsDNA and form complementary pairs with specific nucleotide base pairs in the dsDNA target sequence. Associated with the organic cyclic compounds are aliphatic amino acids, particularly aliphatic amino acids

having a terminal amino group. In addition, a terminus will desirably have a polar group, conveniently substituted on an alkyl substituent. There will be a consecutive series of at least three complementary pairs of organic heterocycles, where by complementary is intended a preferential juxtaposition with a complementary pair of nucleotides. By  
5 appropriate selection of complementary pairs, unpaired organic cyclic compounds in juxtaposition to particular nucleotides of base pairs, aliphatic amino acids, and a polar group substituent, high affinities and high specificities as compared to single-base mismatches can be achieved. The subject compositions are shown to be capable of being transported across cellular membranes to the nucleus, binding to chromosomal DNA, and  
10 fulfilling a variety of intracellular functions, including inhibiting transcription. The compositions may be modified to be used in diagnostics, particularly by providing for detectable labels, or may be used in research or therapeutics, to inhibit transcription of target genes. The compositions may be otherwise modified to enhance properties for specific applications, such as transport across cell walls, association with specific cell  
15 types, cleaving of nucleic acids at specific sites, change chemical and physical characteristics, and the like.

The oligomers of the subject invention will have at least six organic heterocyclic groups, more usually at least seven, and may have eight or more, usually not more than  
20 about thirty, more usually not more than about twenty, frequently not more than about 18, organic cyclic groups, wherein at least 60%, preferably at least 80%, and more preferably at least 100% are heterocycles. The heterocycles generally have from one to three, more usually from one to two heteroatoms, where the heteroatoms are nitrogen, oxygen and sulphur, particularly nitrogen. The nitrogen atoms may be substituted, depending upon  
25 whether the nitrogen atom is directed toward the floor or surface of the groove or away from the groove. Greater latitude in the nature of the substitution is permitted when the nitrogen atom is directed away from the floor of the groove. The orientation of the oligomer is preferably N to C in association with the 5' to 3' direction of the strand to which it is juxtaposed.

30

The heterocycles may be substituted at positions of the heterocycle which are directed away from the floor of the groove for any purpose. Thus, a hydrogen atom may be substituted with a substituent of interest, where the substituent will not result in steric interference with the wall of the minor groove or otherwise create repulsion. When



substituted, the substituents may be widely varied, being heteroatom, hydrocarbyl of from 1 to 30, more usually 1 to 20, carbon atoms, particularly 1 to 10, more particularly 1 to 6 carbon atoms, including aliphatic, alicyclic, aromatic, and combinations thereof, including both aliphatically saturated and unsaturated, having not more than 10% of the carbon

5 atoms participating in aliphatic unsaturation, heterosubstituted hydrocarbyl (as defined previously), having from 1 to 10, usually 1 to 8, more usually 1 to 6, heteroatoms, including aliphatic, alicyclic, aromatic and heterocyclic, and combinations thereof, where the heteroatoms are exemplified by halogen, nitrogen, oxygen, sulfur, phosphorous, metal

10 atoms, boron, arsenic, selenium, rare earths, and the like, wherein functional groups are exemplified by amino, including mono- and disubstituted amino, oxy, including hydroxy and oxyether, thio, including mercapto and thioether, oxo, including oxo-carbonyl (aldehyde and ketone) and non-oxo-carbonyl (carboxy, including acyl halide, anhydride, ester, and amide), phosphorous, including phosphines, phosphites, phosphates, phosphoramidites, etc., boron, including borates, borinic acids and borinates, nitro,

15 cyano, azo, azoxy, hydrazino, etc.

The functional groups may be bonded to an annular member or to a substituent bonded to an annular member, e.g. carboxyalkyl, methoxyethyl, methoxymethyl, aminoethyl, dialkylaminopropyl, polyoxyethylene, polyaminoethylene, etc. In many

20 cases, for annular nitrogen substituents, conveniently, they will be substituted with an alkyl group of from 1 to 3 carbon atoms, particularly methyl, and at least one adjacent annular carbon atom unsubstituted. For the most part, individual substituents will be under 600 Dal, usually under about 300 Dal, and preferably under about 150 Dal and the total for substituents bonded to annular members will be under about 5 kDal, usually under

25 about 2 kDal, more usually under about 1 kDal, there generally being from about 0 to 5, more usually from about 0 to 3 substituents, for other than the alkyl of from 1 to 3 carbon atoms bonded to annular nitrogen. Generally, the total carbon atoms for the substituents will not be greater than about 100, usually not greater than about 60, more usually not greater than about 30, with not more than 30 heteroatoms, usually not more than 20

30 heteroatoms, more usually not more than about 10 heteroatoms.

The heterocycles will normally be linked at the 2 position and the 4 or 5 position, particularly the 2 and 4 position for 5 annular member rings.

The heterocycles are five to six annular members, particularly five annular members, having from one to three, usually one to two heteroatoms, where two heteroatoms are usually spaced apart by at least one intervening carbon atom. The organic cyclic groups are completely unsaturated and will be referred to as aromatic as that term is understood for organic cyclic compounds of from five to six annular members.

Illustrative annular members include pyrrole, imidazole, triazole, furan, thiophene, oxazole, thiazole, pyrazole, cyclopentadiene, pyridine, pyrimidine, triazine, and the like, where as indicated above, NH groups in the rings when substituted are preferably alkylated with an alkyl group of from one to three carbon atoms, particularly methyl. The preferred organic cyclic compounds are five membered rings having from one to two nitrogen atoms, where one of the nitrogen atoms is methylated.

The linking groups between the organic cyclic groups will generally have a length of two atoms, wherein at least some of the linking groups will have NH, where the NH may hydrogen bond with an unshared pair of electrons of the nucleotides. The linking chains may be methyleneamino, carbamyl (-CONH-), ethylene, thiocarbamyl, imidynyl, and the like, particularly carbamyl and its heteroanalogs, e.g. thio and imino.

In addition to the organic cyclic compounds, aliphatic amino acids are employed, particularly  $\omega$ -amino aliphatic amino acids, either to provide for hairpin turns to provide complementation between two sequences of heterocycles, to form a cyclic compound where the oligomers are joined at both ends, or to provide for a shift in spacing of the organic cyclic compounds in relation to the target dsDNA. For the most part, the aliphatic amino acids will have a chain as a core structure of two to six carbon atoms, usually of two to four carbon atoms, desirably having terminal amino groups, particularly glycine,  $\beta$ -alanine, and  $\gamma$ -aminobutyric acid, being unsubstituted or substituted on carbon and nitrogen, particularly carbon, although for the most part the aliphatic amino acids will be unsubstituted. The substituents have been described previously. Where an aliphatic amino acid is C-terminal, the carboxyl group will usually be functionalized as an ester or amide, where the alcohol or amino acid may be selected to provide for specific properties or be used to reduce the charge of the carboxyl group. For the latter, the alcohol and amino groups will generally be from 0 to 6 carbon atoms, usually from 0 to 3 carbon atoms.

As indicated above, these amino acids will play specific roles. The longer chain aliphatic amino acid will serve to provide for turns in the molecule and to close the molecule to form a ring. The shorter chain aliphatic amino acids will be employed, both to provide a shift for spacing in relation to the target dsDNA, and to provide enhanced binding by being present proximal to the terminal organic cyclic group. The aliphatic amino acid may be present at one or both ends of the oligomer. Of particular interest are glycine and alanine, for space-shifting,  $\beta$ -alanine is preferred. Usually, a consecutive sequence of 6 heterocycles will be avoided. Generally, there will be an amino acid, particularly  $\beta$ -alanine, introduced in an otherwise consecutive series of six oligomer units, generally bordered by at least one, preferably at least two organic cyclic groups, particularly heterocycles. The following table indicates the effect of extension of the oligomer heterocycles without introducing an amino acid in the chain.

Table 1\*

	polyamide	sub-unit	binding site size, bp	match	mis-match	specificity
15	Im-(Py) <sub>2</sub> -Dp	3	5	$1.3 \times 10^5$ (0.3)	$< 2 \times 10^4$ <sup>+</sup>	$> 6.5$ <sub>&amp;</sub>
	Im-(Py) <sub>3</sub> -Dp	4	6	$8.5 \times 10^6$ (1.3)	$1.6 \times 10^6$ (0.2)	5.3 (0.5)
	Im-(Py) <sub>4</sub> -Dp	5	7	$4.5 \times 10^7$ (1.1)	$7.9 \times 10^6$ (1.8)	5.7 (0.8)
	Im-(Py) <sub>5</sub> -Dp	6	8	$5.3 \times 10^7$ (0.5)	$< 2 \times 10^7$ <sup>+</sup>	$> 2.7$ <sub>&amp;</sub>
	Im-(Py) <sub>6</sub> -Dp	7	9	$4.7 \times 10^7$ (0.4)	$1.7 \times 10^7$ (0.7)	2.8 (0.7)
20	Im-(Py) <sub>7</sub> -Dp	8	10	$< 2 \times 10^7$	$< 2 \times 10^6$	$\sim 1$

\* Values reported are the mean values from at least three footprint titration experiments. Numbers in parentheses indicate the standard deviation for each data set. The assays were performed at 22°C, pH 7.0, in the presence of 10 mM TrisHCl, 10 mM KCl, 10 mM MgCl<sub>2</sub> and 5 mM CaCl<sub>2</sub>.

- Defined as the ratio of the match site affinity to the affinity of the single base pair mismatch site. Numbers in parentheses indicate the uncertainty calculated using the standard deviations of the measured binding affinities.

& Represents a lower limit on the specificity.

<sup>+</sup> Represents an upper limit for the binding affinity.

The aliphatic chains of the aliphatic amino acids may serve as sites of substitution, the aliphatic amino acid providing a core structure, there usually being not more than 2, more usually not more than 1, substituent. The same types of substituents that have been

described for the heterocycles may also be employed here. Conveniently, the substituted aliphatic amino acid may be used in the synthesis of the oligomer, rather than modifying the amino acid after the oligomer is formed. Alternatively, a functional group may be present on the chain of the substituent, if necessary being appropriately protected during the course of the synthesis, which functional group may then be used for the subsequent modification. Desirably, such functional group could be selectively used, for synthesis of different oligomers, so as to provide for substitution at that site to produce products having unique properties associated with a particular application. With the substituent substituted at a site which does not significantly interfere with the binding in the groove, e.g. employing a single stereoisomer, properties can be imparted to the subject compounds, such as water solubility, lipophilicity, non-covalent binding to a receptor, radioactivity, fluorescence, etc.

One or both termini, preferably one of the termini, will have a polar group substituted on an alkyl group, where the polar group will generally be from 2 to 6, more usually 2 to 4, carbon atoms from the linkage to the remaining molecule. The polar group may be charged or uncharged, where the charge may be a result of protonation under the conditions of use. Particularly, groups capable of hydrogen bonding are preferred, such as amino, particularly ~~tertiary~~-amino, hydroxyl, mercapto, and the like. Of particular interest is amino, more particularly alkylated amino, where the alkyl groups are of from 1 to 6, usually 1 to 3, more usually 1, carbon atom, and at a pH less than about eight, the amino group is positively charged, and can hydrogen bond with the dsDNA. Desirably, two positively charged polar groups will not be employed on the oligomers, where the positively charged polar groups will be in juxtaposition when complexed with the dsDNA.

For many purposes one may wish to have an isotopic oligomer, where one can analyze for its presence, using scintillation counters for radioactive elements, nmr for atoms having a magnetic moment, and the like. For a radioactive oligomer, a radioactive label may be employed, such as tritium,  $^{14}\text{C}$ ,  $^{125}\text{I}$ , or the like. The radiolabel may be a substituent on an annular member of a heterocycle or an annular member of a heterocycle, either carbon or a heteroatom, or a substituent at the C- or N-terminus of the oligomer, depending upon convenience. By using a radiolabel as part of the oligomer, one avoids any significant change in the spatial conformation of the oligomer. The radiolabel may serve numerous purposes in diagnostics, cytohistology, radiotherapy, and the like.

Besides the other sites present on the oligomer, either terminus of the oligomer may be used for special purposes depending upon the use to which the oligomer is put. For example, in diagnostics, one may wish to have a detectable label other than a radiolabel, where the resulting compound may find use for other purposes, as well. The

5 oligomer may be linked to labels, such as fluorescers, e.g. dansyl, fluorescein, Texas red, isosulfan blue, ethyl red, malachite green, etc., chemilumescers, particles, e.g. magnetic particles, colloidal particles, e.g. gold particles, light sensitive bond forming compounds, e.g. psoralens, anthranilic acid, pyrene, anthracene, and acridine, chelating compounds, such as EDTA, NTA, tartaric acid, ascorbic acid, polyhistidines of from 2 to 8 histidines,

10 alkylene polyamines, etc., chelating antibiotics, such as bleomycin, where the chelating compounds may chelate a metal atom, such as iron, cobalt, nickel, technetium, etc., where the metal atom may serve to cleave DNA in the presence of a source of peroxide, intercalating dyes, such as ethidium bromide, thiazole orange, thiazole blue, TOTO, 4',6-diamidino-2-phenylindole (DAPI), etc., enzymes, such as  $\beta$ -galactosidase, NADH or

15 NADHP dehydrogenase, malate dehydrogenase, lysozyme, peroxidase, luciferase, etc., alkylating agents such as haloacetamides, N-ethyl nitrosourea, nitrogen and sulfur mustards, sulfonate esters, etc., and other compounds, such as arylboronic acids, tocopherols, lipoic acid, captothesin, etc. colloidal particles, e.g. gold particles, fluorescent particles, peroxides, DNA cleaving agents, oligonucleotides, oligopeptides,

20 nmr agents, stable free radicals, metal atoms, etc. The oligomer may be combined with other labels, such as haptens for which a convenient receptor exists, e.g. biotin, which may be complexed with avidin or streptavidin and digoxin, which may be complexed with antidigoxin, etc. where the receptor may be conjugated with a wide variety of labels, such as those described above. The oligomers may be joined to sulfonated or phosphonated

25 aromatic groups, e.g. naphthalene, to enhance inhibition of transcription, particularly of viruses (Clanton et al., Antiviral Res. (1995) 27:335 -354). In some instances, one may bond multiple copies of the subject oligomers to polymers, where the subject oligomers are pendant from the polymer. Polymers, particularly water soluble polymers, which may find use are cellulose, poly(vinyl alcohol), poly(vinyl acetate-vinyl alcohol), polyacrylates,

30 and the like. The number of oligomers may be from 1 to about 1:5 monomer units of the polymer.

One may wish to enhance the lipophilicity of the molecule, providing for various lipophilic groups, such as cholesterol, fatty acids, fatty alcohols, sphingomyelins,

cerebrosides, other glycerides, and the like, where the fatty group will generally be of from about eight to thirty carbon atoms. Alternatively, one may wish to provide for saccharides, which bind to lectins, adhesion molecules, bacteria, or the like, where the saccharides serve to direct the subject oligomers to a specific cellular target.

- 5 Alternatively, in some instances, one may wish to have one or more nucleotides, generally from about one to thirty, more usually from about three to twenty, particularly from about three to twelve. The nucleotides will normally be associated with the proximal or bordering nucleic acid sequence of the target sequence, whereby the attached nucleic acid sequence will complex with the nucleotides in the major groove.

10

- The different molecules may be joined to the termini in a variety of ways, depending upon the available functionality(ies) present at the termini, such as extending the polar substituted alkyl group, e.g. having a chain of more than 6 carbon atoms, providing for a substituent at a terminus which can be reacted with the moiety to be added,
- 15 where such substituents will conventionally be amino, hydroxyl, mercapto, carboxyl, phosphate, etc., so as to form amides, both organic and inorganic, substituted amines (reductive amination), ethers, thioethers, disulfides, esters, both organic and inorganic, pyrophosphates, and the like. The molecules may be introduced as part of the synthetic scheme, displacing the oligomer from the solid support on which the oligomer is
- 20 synthesized. Because the compounds of the subject invention may be used in such a variety of ways, no simple description is appropriate to the variety of moieties to which the subject oligomers may be bound, nor the specific molecular weights of the resulting products.

- 25 The subject oligomers may be synthesized on supports, e.g. chips, where by using automated synthetic techniques, different oligomers may be synthesized at individual sites. In this way, an array of different oligomers may be synthesized, which can then be used to identify the presence of a plurality of different sequences in a sample. By knowing the composition of the oligomer at each site, one can identify binding of specific sequences at
- 30 that site by various techniques, such as labeled antiDNA antibodies, linkers having complementary restriction overhangs, where the sample DNA has been digested with a restriction enzyme, and the like. The techniques for preparing the subject arrays are analogous to the techniques used for preparing oligopeptide arrays, as described in Cho et al., Science, 1993, 261, 1303-1305.

The complex will usually comprise one or two oligomers or combinations of one or two oligomers, where individual or pairs of oligomers specifically interact with a dsDNA sequence of at least 6, usually at least 7 and preferably 8 or more bp, frequently not more than 40 bp, more usually not more than about 30 bp, preferably not more than 20  
5 bp.

Since a major portion of the work has been performed with N-methyl pyrrole and N-methyl imidazole, using carbamyl groups as the linking chains, with the aliphatic amino acids glycine,  $\beta$ -alanine and  $\gamma$ -aminobutyric acid, as well as dimethylaminopropyl as the  
10 polar substituted alkyl group, these compounds will now be illustrated as exemplary of the class of compounds which may be employed in the subject invention. It is understood that one or a few of the nitrogen-heterocycles may be substituted with a different organic cyclic group, as well one or the other of the aliphatic amino acids may be substituted with a different amino acid, etc. Furthermore, the core oligomer may be further substituted for  
15 specific applications as described above. In effect, there is a core molecule or core molecules which define at least complementary pairs of heterocycles, and include at least one of an aliphatic amino acid and a polar group substituted alkyl. This core molecule which is the centerpiece of the invention can serve as the nexus for numerous substitutions which do not interfere with the basic function of the core molecule, although where the  
20 binding affinity is greater than is necessary for the function, some degradation of the binding affinity is permitted. Therefore, in defining the compounds of this invention, it should be understood that many variations are permitted, where the basic core or unit structure is retained, while the core or unit structure is modified with one or more substituents to impart desired properties to the molecule for its intended function.

25 Of particular interest among the subject compounds are compounds which have at least one organic cyclic group, particularly N-methyl imidazole, which has specificity for one nucleotide, which is present as a complementary pair. Usually, the subject compounds will have at least one of these complementary pairs, frequently at least two of  
30 these complementary pairs, and generally fewer than 75% of the complementary pairs will have the organic cyclic group having specificity for a single nucleotide. In the case of the N-methyl imidazole, there will usually be at least one Im/Py pair, desirably not having more than three, frequently having not more than two, of such pairs consecutively, so that there will frequently be not more than three Im's in a row. There will normally be at least

one aliphatic amino acid, frequently two aliphatic amino acids, and frequently not more than eight aliphatic amino acids, usually not more than six aliphatic amino acids, more usually not more than about four aliphatic amino acids. Preferably, there will be an amino acid proximal to at least one terminus of the oligomer. The Im/Py pair provide for greater  
5 specificity, and when appropriately placed contribute in at least a similar manner to the Py/Py pair to the binding affinity for dsDNA. Therefore, by appropriate selection of the target sequence, one may optimize for binding affinity and specificity.

It is found that with  $\beta$ -alanine,  $\beta$ -alanine associates with T-A pairs and will  
10 usually form a complementary pair with itself. Thus,  $\beta$ -alanine may be used in juxtaposition to T or A and as a complementary pair with itself with a T-A pair.

The binding affinity  $K_a$  as determined in the Experimental section will be greater than  $5 \times 10^8 \text{ M}^{-1}$ , usually greater than  $10^9 \text{ M}^{-1}$ , preferably greater than about  $10^{10} \text{ M}^{-1}$ , so as  
15 to be able to bind to the target sequence at subnanomolar concentrations in the environment in which they are used. The difference in affinity with a single mismatch will be at least 3 fold, usually at least 5 fold, preferably at least 10 fold and frequently greater than 20 fold, and may be 100 fold or more.

20 Where the oligomers of the subject invention are used with cells, particularly viable cells, the oligomers will generally have a molecular weight of less than about 5kD, preferably less than about 3.5kD, and will generally have a molecular weight of at least about .6kD, more usually at least about .8kD.

25 The compositions of the subject invention for complexing with dsDNA will have from one to two oligomers, or combinations thereof, depending upon whether there is a hairpin turn in the oligomer, where only one oligomer is necessary, or there is no hairpin turn, so that for complementarity, one needs two oligomers. More oligomers may be used, where one wishes to target more than one dsDNA sequence, for example,  
30 contiguous or proximal sequences, to enhance the overall specificity, or for distal sequences, where the sequences may be associated with the same functional unit, e.g. a gene, or different functional units, e.g. homeodomains. The composition, whether a single oligomer or a combination of oligomers will provide at least three complementary pairs in the single oligomer or pair of oligomers.



In many cases, in order to achieve the desired association constants, one will increase the number of complementary pairs and/or have regions of unpaired organic cyclic groups. Usually, one will have at least one or both of a fourth complementary pair or at least two unpaired organic cyclic groups, so as to have a chain of four organic cyclic groups involved in pair formation and/or at least two organic cyclic groups uninvolved with pair formation. It is found that one does not increase the binding affinity to the same extent with each addition of an organic cyclic unit, as one extends the length of the oligomer and, in fact, as described previously, one may begin to reverse the binding affinity by the continuous extension. Therefore, by appropriate choice, as indicated above, one can limit the composition and size of individual oligomers to optimize the binding affinity, as well as the other properties which are associated with the oligomeric composition.

Because of the extensive utilization of N-methyl pyrrole and N-methyl imidazole, the following compounds which employ these N-heterocycles are exemplary of the class of compounds of the subject invention. These compounds may be prepared in accordance with the procedures described herein. When used, Py will refer to N-methyl pyrrole and Im will refer to N-methyl imidazole.

ImPyPyPy- $\gamma$ -PyPyPyPy, PyPyImPy- $\gamma$ -PyPyPyPy, ImPyPyPy- $\gamma$ -ImPyPyPy, PyImPyPy- $\gamma$ -PyImPyPy, ImPyImPy- $\gamma$ -PyPyPyPy, ImImPyPy- $\gamma$ -PyPyPyPy, ImImImPy- $\gamma$ -PyPyPyPy, ImImPyPy- $\gamma$ -ImPyPyPy, ImPyPyPy- $\gamma$ -ImImPyPy, ImImPyPy- $\gamma$ -ImImPyPy, ImPyImPy- $\gamma$ -ImPyImPy, ImImImPy- $\gamma$ -ImPyPyPyPy, ImImImIm- $\gamma$ -PyPyPyPy, Im- $\beta$ -PyPy- $\gamma$ -Im- $\beta$ -PyPy, Im- $\beta$ -ImIm- $\gamma$ -Py- $\beta$ -PyPy, Im- $\beta$ -ImPy- $\gamma$ -Im- $\beta$ -ImPy, ImPyPyPyPy- $\gamma$ -ImPyPyPyPy, ImImPyPyPy- $\gamma$ -ImPyPyPyPy, ImPyImPyPy- $\gamma$ -ImPyPyPyPy, ImImPyImIm- $\gamma$ -PyPyPyPyPy, ImPyPyImPy- $\gamma$ -ImPyPyImPy, ImPy- $\beta$ -PyPy- $\gamma$ -ImPy- $\beta$ -PyPy, ImIm- $\beta$ -ImIm- $\gamma$ -PyPy- $\beta$ -PyPy, ImPy- $\beta$ -ImPy- $\gamma$ -ImPy- $\beta$ -ImPy, ImPy- $\beta$ -PyPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPy, ImIm- $\beta$ -PyPyPy- $\gamma$ -PyPyPy- $\beta$ -PyPy, ImPy- $\beta$ -ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPy, ImIm- $\beta$ -PyPyPy- $\gamma$ -ImImPy- $\beta$ -PyPy, ImPy- $\beta$ -PyPyPy- $\gamma$ -PyPyPy- $\beta$ -ImPy, ImPyPyPyPyPy- $\gamma$ -ImPyPyPyPyPy, ImPyPy- $\beta$ -PyPy- $\gamma$ -ImPyPy- $\beta$ -PyPy, ImPyPyPy- $\beta$ -Py- $\gamma$ -Im- $\beta$ -PyPyPyPy, ImImPyPyPyPy- $\gamma$ -ImImPyPyPyPy, Im- $\beta$ -PyPyPyPy- $\gamma$ -Im- $\beta$ -PyPyPyPy, ImPyPyPy- $\beta$ -Py- $\gamma$ -ImPyPyPy- $\beta$ -Py, ImPyImPyPyPy- $\gamma$ -ImPyPyPyPyPy, ImPyPy- $\beta$ -PyPy- $\gamma$ -ImPy- $\beta$ -PyPyPy, ImPyPyPyPy- $\beta$ - $\gamma$ -ImPyPyPyPy- $\beta$ , ImPy- $\beta$ -ImPyPy- $\gamma$ -ImPy- $\beta$ -ImPyPy, Im- $\beta$ -PyPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Py, Im- $\beta$ -

ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Py, ImPyPy- $\beta$ -PyPyPy, ImImPy- $\beta$ -PyPyPy, ImImIm- $\beta$ -  
 PyPyPy, ImPyPyPyPy- $\beta$ -PyPyPy, ImPyPyPy- $\beta$ -PyPyPy, ImPyPy- $\beta$ -PyPyPyPyPy,  
 ImPyPyPy- $\beta$ -PyPyPyPy, ImImPyPy- $\beta$ -PyPyPyPy, ImImImPy- $\beta$ -PyPyPyPy, ImPyPyPy- $\beta$ -  
 ImPyPyPy, ImImPyPy- $\beta$ -ImPyPyPy, ImImPyPyPy- $\beta$ -PyPyPyPyPy, ImImImPyPy- $\beta$ -  
 5 PyPyPyPyPy, ImIm- $\beta$ -PyPy- $\beta$ -PyPy- $\beta$ -PyPy, ImImPy- $\beta$ -PyPyPy- $\beta$ -PyPyPy, ImImPyPy-  
 $\beta$ -Py- $\beta$ -PyPyPyPy, ImPyPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy, ImPyPy- $\gamma$ -PyPyPy- $\beta$ -PyPyPy,  
 PyImPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy, PyImPy- $\gamma$ -ImPyPy- $\beta$ -PyPyPy- $\beta$ -PyPyPy, ImImPy- $\gamma$ -  
 ImPyPy- $\beta$ -PyPyPy, ImPyPy- $\gamma$ -ImPyPy-G-PyPyPy, ImPyPyPy- $\gamma$ -ImImImPy- $\beta$ -PyPyPyPy,  
 ImImPyPy- $\gamma$ -ImImPyPy- $\beta$ -PyPyPyPy, and ImImPyPy- $\gamma$ -PyPyPyPy- $\beta$ -PyPyPyPy,  
 10 PyPyImIm, ImPy- $\beta$ -ImPy- $\beta$ -ImPy, ImPy- $\beta$ -ImPy- $\beta$ -ImPy- $\beta$ -PyPy, ImPy- $\beta$ -ImPy- $\beta$ -ImPy-  
 $\gamma$ -ImPy- $\beta$ -ImPy- $\beta$ -ImPy, ImPy- $\beta$ -PyPy- $\beta$ -PyPy, ImPy- $\beta$ -PyPy- $\beta$ -PyPy- $\gamma$ -ImPy- $\beta$ -PyPy-  
 $\beta$ -PyPy, Im- $\beta$ -ImImImIm- $\gamma$ -PyPyPyPy- $\beta$ -Py, Im- $\beta$ -ImImImIm- $\beta$ -Im- $\gamma$ -Py- $\beta$ -PyPyPyPy- $\beta$ -  
 Py, ImIm- $\beta$ -Im- $\gamma$ -Py- $\beta$ -PyPyPyPy- $\beta$ -Py, Im- $\beta$ -ImPyPy- $\gamma$ -ImImPy- $\beta$ -Py, ImImPy- $\beta$ -Py- $\gamma$ -  
 Im- $\beta$ -ImPyPy, ImIm- $\beta$ -Im- $\gamma$ -PyImPyPy.

15

Figure 1 illustrates the relationship between the azoles and the nucleotides of the minor groove.

Where two oligomers are used, the oligomers may be completely overlapped, or  
 20 only partially overlapped, i.e. slipped or have overhangs. As indicated previously, there  
 will be at least 3 complementary azole (N-methyl pyrrole and imidazole)pairs. In the  
 overlapped configuration, all of the azoles are in complementary pairs, as well as any  
 spacing amino acid. In the slipped configuration, there will be at least one azole ring  
 which is unpaired in at least one of the oligomers, usually there will be at least two azole  
 25 rings, more usually, in both of the oligomers. Usually, the number of unpaired azole  
 rings will be in the range of 2 to 30, more usually 2 to 20, frequently 2 to 12. Generally,  
 unpaired azoles will involve chains of 2 or more azole rings, more usually 3 or more azole  
 rings, including, as appropriate, aliphatic amino acids in the chain.

30 Various permutations and combinations of oligomers may be used. One may  
 have a single oligomer having at least three complementary pairs and an extension or  
 overhang of unpaired azoles, which may be complemented in whole or part by a second  
 oligomer, which forms complementary pairs with the unpaired members of the first  
 oligomer. Alternatively, one may have two "candy cane" oligomers, having

complementary pairs, with the members of the complementary pairs separated by a  $\gamma$ -aminobutyric acid, and an overhang of unpaired members. However, these otherwise unpaired members of one oligomer can be positioned to form complementary pairs with the extension or overhang of the unpaired members of the other oligomer. One may have  
5 an extended linear oligomer, where two or more oligomers complement the azoles of the extended linear oligomer. If one wished, one could have alternating regions of unpaired and paired azoles by using a plurality of oligomers which complement to various degrees. In each case, the selection would be related to the desired affinity, the nature of the target, the purpose for the formation of the complex, and the like.

10

The subject compositions may be brought together with the dsDNA under a variety of conditions. The conditions may be in vitro, in cell cultures, ex vivo or in vivo. For detecting the presence of a target sequence, the dsDNA may be extracellular or intracellular. When extracellular, the dsDNA may be in solution, in a gel, on a slide, or  
15 the like. The dsDNA may be present as part of a whole chromosome or fragment thereof of one or more centiMorgans. The dsDNA may be part of an episomal element. The dsDNA may be present as smaller fragments ranging from about 20, usually at least about 50, to a million base pairs, or more. The dsDNA may be intracellular, chromosomal, mitochondrial, plastid, kinetoplastid, or the like, part of a lysate, a chromosomal spread,  
20 fractionated in gel electrophoresis, a plasmid, or the like, being an intact or fragmented moiety. The formation of complexes between dsDNA and the subject compounds may be for diagnostic, therapeutic, purification, or research purposes, and the like. Because of the specificity of the subject compounds, the subject compounds can be used to detect specific dsDNA sequences in a sample without melting of the dsDNA. The diagnostic  
25 purpose for the complex formation may be detection of alleles, identification of mutations, identification of a particular host, e.g. bacterial strain or virus, identification of the presence of a particular DNA rearrangement, identification of the presence of a particular gene, e.g. multiple resistance gene, forensic medicine, or the like. With pathogens, the pathogens may be viruses, bacteria, fungi, protista, chlamydia, or the like. With higher  
30 hosts, the hosts may be vertebrates or invertebrates, including insects, fish, birds, mammals, and the like or members of the plant kingdom.

When involved in vitro or ex vivo, the dsDNA may be combined with the subject compositions in appropriately buffered medium, generally at a concentration in the range

of about 0.1 nM to 1 mM. Various buffers may be employed, such as TRIS, HEPES, phosphate, carbonate, or the like, the particular buffer not being critical to this invention. Generally, conventional concentrations of buffer will be employed, usually in the range of about 10-200mM. Other additives which may be present in conventional amounts include  
5 sodium chloride, generally from about 1-250 mM, dithiothreitol, and the like, the particular nature of quantity of salt not being critical to this invention. The pH will generally be in the range of about 6.5 to 9, the particular pH not being critical to this invention. The temperature will generally be in a range of 4°C to 45°C, the particular temperature not being critical to this invention. The target dsDNA may be present in from  
10 about 0.001 to 100 times the moles of oligomer.

The subject compounds when used in diagnosis may have a variety of labels as indicated previously and may use many of the protocols that have been used for detection of haptens and receptors (immunoassays) or with hybridization (DNA complementation).  
15 Since the subject compounds are not nucleic acids, they can be employed more flexibly than when using DNA complementation. The assays are carried out as described below and then depending on the nature of the label and protocol, the determination of the presence and amount of the sequence may then be made. The protocols may be performed in solution or in association with a solid phase. The solid phase may be a vessel wall, a  
20 particle, fiber, film, sheet, or the like, where the solid phase may be comprised of a wide variety of materials, including gels, paper, glass, plastic, metals, ceramics, etc. Either the sample or the subject compounds may be affixed to the solid phase in accordance with known techniques. By appropriate functionalization of the subject compounds and the solid phase, the subject compounds may be covalently bound to the solid phase. The  
25 sample may be covalently or non-covalently bound to the solid phase, in accordance with the nature of the solid phase. The solid phase allows for a separation step, which allows for detection of the signal from the label in the absence of unbound label.

Exemplary protocols include combining a cellular lysate, with the DNA bound to  
30 the surface of a solid phase, with an enzyme labeled oligomer, incubating for sufficient time under complex forming conditions for the oligomer to bind to any target sequence present on the solid phase, separating the liquid medium and washing, and then detecting the presence of the enzyme on the solid phase by use of a detectable substrate.

A number of protocols are based on having a label which does not give a detectable signal directly, but relies on non-covalent binding with a receptor, which is bound to a surface or labeled with a directly detectable label. In one assay one could have a hapten, e.g. digoxin, bonded to the oligomer. The sample DNA would be bound to a surface, so as to remain bound to the surface during the assay process. The oligomer would be added and bind to any target sequence present. After washing to remove oligomer, enzyme or fluorescer labeled antidigoxin monoclonal antibody is added, the surface washed and the label detected. Alternatively, one may have a fluorescer bound to one end of the oligomer and biotin or other appropriate hapten bound to the other end of the oligomer or to the complementary oligomer. The oligomers are combined with the DNA in the liquid phase and incubated. After completion of the incubation, the sample is combined with the receptor for the biotin or hapten, e.g. avidin or antibody, bound to a solid surface. After a second incubation, the surface is washed and the level of fluorescence determined.

15

If one wishes to avoid a separation step, one may use channeling or fluorescence quenching. By having two labels which interact, for example, two enzymes, where the product of one enzyme is the substrate of the other enzyme, or two fluorescers, where there can be energy transfer between the two fluorescers, one can determine when complex formation occurs, since the two labels will be brought into juxtaposition by forming the 2:1 complex in the minor groove. With the two enzymes, one detects the product of the second enzyme and with the two fluorescers, one can determine fluorescence at the wavelength of the Stokes shift or reduction in fluorescence of the fluorescer absorbing light at the lower wavelength. Another protocol would provide for binding the subject compositions to a solid phase and combining the bound oligomers with DNA in solution. After the necessary incubations and washings, one could add labeled antiDNA to the solid phase and determine the amount of label bound to the solid phase.

To determine a number of different sequences simultaneously or just a single sequence, one may provide an array of the subject compositions bound to a surface. In this way specific sites in the array will be associated with specific DNA sequences. One adds the DNA containing sample to the array and incubates. DNA which contains the complementary sequence to the oligomer at a particular site will bind to the oligomers at that site. After washing, one then detects the presence of DNA at particular sites, e.g.

with an antiDNA antibody, indicating the presence of the target sequence. By cleaving the DNA with a restriction enzyme in the presence of a large amount of labeled linker, followed by inactivation of the enzyme, one may then ligate the linker to the termini of the DNA fragments and proceed as described above. The presence of the label at a particular  
5 site in the array will indicate the presence of the target sequence for that site.

The number of protocols that may be used is legion. Illustrative protocols may be found for DNA assays in WO95/20591; EPA 393,743; WO86/05519; and EPA 278,220, while protocols and labels which may be adapted from immunoassays for use with the  
10 subject compositions for assays for DNA may be found in WO96/20218; WO95/06115; WO94/04538; WO94/01776; WO92/14490; EPA 537830; WO91/09141; WO91/06857; and WO91/05257.

During diagnostics, such as involved with cells, one may need to remove the non-  
15 specifically bound oligomers. This can be achieved by combining the cells with a substantial excess of the target sequence, conveniently attached to particles. By allowing for the non-specifically bound oligomers to move to the extracellular medium, the oligomers will become bound to the particles, which may then be readily removed. If desired, one may take samples of cells over time and plot the rate of change of loss of the  
20 label with time. Once the amount of label becomes stabilized, one can relate this value to the presence of the target sequence. Other techniques may also be used to reduce false positive results.

The subject compositions may also be used to titrate repeats, where there is a  
25 substantial change, increase or decrease, in the number of repeats associated with a particular indication. The number of repeats should be at least an increase of 50%, preferably at least two-fold, more preferably at least three-fold. By determining the number of oligomers which become bound to the dsDNA, one can determine the amplification or loss of a particular repeat sequence.

30

The subject compositions may be used for isolation and/or purification of target DNA comprising the target sequence. By using the subject oligomers, where the oligomers are bound to a solid phase, those portions of a DNA sample which have the target sequence will be bound to the subject oligomers and be separated from the

remaining DNA. One can prepare columns of particles to which the oligomers are attached and pass the sample through the column. After washing the column, one can release the DNA which is specifically bound to the column using solvents or high salt solutions. Alternatively, one can mix particles to which the oligomers are bound with the sample and then separate the particles, for example, with magnetic particles, using a magnetic field, with non-magnetic particles, using centrifugation. In this way, one can rapidly isolate a target DNA sequence of interest, for example, a gene comprising an expressed sequence tag (EST), a transcription regulatory sequence to which a transcription factor binds, a gene for which a fragment is known, and the like. As partial sequences are defined by a variety of techniques, the subject oligomers allow for isolation of restriction fragments, which can be separated on a gel and then sequenced. In this way the gene may be rapidly isolated and its sequence determined. As will be discussed below, the subject oligomers may then be used to define the function of the gene.

The subject oligomers may be used in a variety of ways in research. Since the subject oligomers can be used to inhibit transcription, the effect of inhibiting transcription on cells, cell assemblies and whole organisms may be investigated. For example, the subject compositions may be used in conjunction with egg cells, fertilized egg cells or blastocysts, to inhibit transcription and expression of particular genes associated with development of the fetus, so that one can identify the effect of reduction in expression of the particular gene. Where the gene may be involved in regulation of a number of other genes, one can define the effect of the absence of such gene on various aspects of the development of the fetus. The subject oligomers can be designed to bind to homeodomains, so that the transcription of one or more genes may be inhibited. In addition, one can use the subject compositions during various periods during the development of the fetus to identify whether the gene is being expressed and what the effect is of the gene at the particular stage of development.

With single cell organisms, one can determine the effect of the lack of a particular expression product on the virulence of the organism, the development of the organism, the proliferation of the organism, and the like. In this way, one can determine targets for drugs to inhibit the growth and infectiousness of the organism.

In an animal model, one can provide for inhibition of expression of particular genes, reversibly or irreversibly, by administering the subject compositions to the host in a variety of ways, oral or parenteral, by injection, at a particular site where one wishes to influence the transcription, intravascularly, subcutaneously, or the like. By inhibiting transcription, one can provide for a reversible "knock out," where by providing for continuous intravenous administration, one can greatly extend the period in which the transcription of the gene is inhibited. Alternatively, one may use a bolus of the subject oligomers and watch the effect on various physiological parameters as the bolus becomes dissipated. One can monitor the decay of the effect of the inhibition, gaining insight into the length of time the effect lasts, the physiological processes involved with the inhibition and the rate at which the normal physiological response occurs. Instead, one can provide for covalent bonding of the oligomer to the target site, using alkylating agents, light activated bonding groups, intercalating groups, etc.

It is also possible to upregulate genes, by downregulating other genes. In those instances where one expression product inhibits the expression of another expression product, by inhibiting the expression of the first product, one can enhance the expression of the second product. Similarly, transcription factors involve a variety of cofactors to form a complex, one can enhance complex formation with one transcription factor, as against another transcription factor, by inhibiting expression of the other transcription factor. In this way one can change the nature of the proteins being expressed, by changing the regulatory environment in the cell.

The target sequence may be associated with the 5'-untranslated region, namely the transcriptional initiation region, an enhancer, which may be in the 5'-untranslated region, the coding sequence or introns, the coding region, including introns and exons, the 3'-untranslated region, or distal from the gene.

The subject compositions may be presented as liposomes, being present of the lumen of the liposome, where the liposome may be combined with antibodies to surface membrane proteins or basement membrane proteins, ligands for cellular receptors, or other site directing compound, to localize the subject compositions to a particular target. See, for example, Theresa and Mouse, Adv. Drug Delivery Rev. 1993, 21, 117-133; Huwlyer and Partridge, Proc. Natl. Acad. Sci. USA 1996, 93, 11421-11425; Dzau et al.,



Proc. Natl. Acad. Sci. USA 1996, 93, 11421-11425; and Zhu et al., Science, 1993, 261, 209-211. The subject compositions may be administered by catheter to localize the subject compositions to a particular organ or target site in the host. Generally, the concentration at the site of interest should be at least about 0.1nM, e.g. intracellular or in the  
5 extracellular medium, preferably at least about 1nM, usually not exceeding 1mM, more usually not exceeding about 100nM. To achieve the desired intracellular concentration, the concentration of the oligomers extracellularly will generally be greater than the desired intracellular concentration, ranging from about 2 to 1000 times or greater the desired intracellular concentration. Of course, where the toxicity profile allows for higher  
10 concentrations than those indicated for intracellular or extracellular concentrations, the higher concentrations may be employed, and similarly, where the affinities are high enough, and the effect can be achieved with lower concentrations, the lower concentrations may also be employed.

15 The subject compositions can be used to modulate physiological processes in vivo for a variety of reasons. In non-primates, particularly domestic animals, in animal husbandry and breeding, one can affect the development of the animal by controlling the expression of particular genes, modify physiological processes, such as accumulation of fat, growth, response to stimuli, etc. One can also use the subject compositions for  
20 therapeutic purposes in mammals. Domestic animals include feline, murine, canine, lagomorpha, bovine, ovine, canine, porcine, etc.

The subject compositions may used therapeutically to inhibit proliferation of particular target cells, inhibit the expression of one or more genes related to an indication;  
25 change the phenotype of cells, either endogenous or exogenous to the host, where the native phenotype is detrimental to the host. Thus, by providing for binding to housekeeping or other genes of bacteria or other pathogen, particularly genes specific to the pathogen, one can provide for inhibition of proliferation of the particular pathogen. Various techniques may be used to enhance transport across the bacterial wall, such as  
30 various carriers or sequences, such as polylysine, poly(E-K), nuclear localization signal, cholesterol and cholesterol derivatives, liposomes, protamine, lipid anchored polyethylene glycol, phosphatides, such as dioleoxyposphatidylethanolamine, phosphatidyl choline, phosphatidylglycerol,  $\alpha$ -tocopherol, cyclosporin, etc. In many cases, the subject compositions may be mixed with the carrier to form a dispersed composition and used as

the dispersed composition. Similarly, where a gene may be essential to proliferation or protect a cell from apoptosis, where such cell has undesired proliferation, the subject compositions can be used to inhibit the proliferation by inhibiting transcription of essential genes. This may find application in situations such as cancers, such as sarcomas, carcinomas and leukemias, restenosis, psoriasis, lymphopoiesis, atherosclerosis, pulmonary fibrosis, primary pulmonary hypertension, neurofibromatosis, acoustic neuroma, tuberous sclerosis, keloid, fibrocystic breast, polycystic ovary and kidney, scleroderma, rheumatoid arthritis, ankylosing spondylitis, myelodysplasia, cirrhosis, esophageal stricture, sclerosing cholangitis, retroperitoneal fibrosis, etc. Inhibition may be associated with one or more specific growth factors, such as the families of platelet-derived growth factors, epidermal growth factors, transforming growth factor, nerve growth factor, fibroblast growth factors, e.g. basic and acidic, keratinocyte fibroblast growth factor, tumor necrosis factors, interleukins, particularly interleukin 1, interferons, etc. In other situations, one may wish to inhibit a specific gene which is associated with a disease state, such as mutant receptors associated with cancer, inhibition of the arachidonic cascade, inhibition of expression of various oncogenes, including transcription factors, such as ras, myb, myc, sis, src, yes, fps/fes, erbA, erbB, ski, jun, crk, sea, rel, fms, abl, met, trk, mos, Rb-1, etc. Other conditions of interest for treatment with the subject compositions include inflammatory responses, skin graft rejection, allergic response, psychosis, sleep regulation, immune response, mucosal ulceration, withdrawal symptoms associated with termination of substance use, pathogenesis of liver injury, cardiovascular processes, neuronal processes. Particularly, where specific T-cell receptors are associated with autoimmune diseases, such as multiple sclerosis, diabetes, lupus erythematosus, myasthenia gravis, Hashimoto's disease, cytopenia, rheumatoid arthritis, etc., the expression of the undesired T-cell receptors may be diminished, so as to inhibit the activity of the T-cells. In cases of reperfusion injury or other inflammatory insult, one may provide for inhibition of enzymes associated with the production of various factors associated with the inflammatory state and/or septic shock, such as TNF, enzymes which produce singlet oxygen, such as peroxidases and superoxide dismutase, proteases, such as elastase,  $\text{INF}\gamma$ , IL-2, factors which induce proliferation of mast cells, eosinophils,  $\text{IgG}_1$ , IgE, regulatory T cells, etc., or modulate expression of adhesion molecules in leukocytes and endothelial cells.

Other opportunities for use of the subject compositions include modulating levels of receptors, production of ligands, production of enzymes, production of factors, reducing specific cell populations, changing phenotype and genotype of cells, particularly as associated with particular organs and tissues, modifying the response of cells to drugs  
5 or other stimuli, e.g. enhancing or diminishing the response, inhibiting one of two or more alleles, repressing expression of target genes, particularly as related to clinical studies, modification of behavior, modification of susceptibility to disease, response to stimuli, response to pathogens, response to drugs, therapeutic or substances of abuse, etc.

10 Individual compositions may be employed or combinations, directed to the same dsDNA region, but different target sequences, contiguous or distal, or different DNA regions. Depending upon the number of genes which one wishes to target, the composition may have one or a plurality of oligomers or pairs of oligomers which will be directed to different target sites.

15

The subject compositions may be used as a sole therapeutic agent or in combination with other therapeutic agents. Depending upon the particular indication, other drugs may also be used, such as antibiotics, antisera, monoclonal antibodies, cytokines, anti-inflammatory drugs, and the like. The subject compositions may be used  
20 for acute situations or in chronic situations, where a particular regimen is devised for the treatment of the patient. The compositions may be prepared in physiologically acceptable media and stored under conditions appropriate for their stability. They may be prepared as powders, solutions or dispersions, in aqueous media, alkanols, e.g. ethanol and propylene glycol, in conjunction with various excipients, etc. The particular formulation will depend  
25 upon the manner of administration, the desired concentration, ease of administration, storage stability, and the like. The concentration in the formulation will depend upon the number of doses to be administered, the activity of the oligomers, the concentration required as a therapeutic dosage, and the like. The subject compositions may be administered orally, parenterally, e.g. intravenously, subcutaneously, intraperitoneally,  
30 transdermally, etc. The subject compositions may be formulated in accordance with conventional ways, associated with the mode of treatment. As a result of the formulation, the subject compositions are introduced into the cells, either as a directed introduction to a specific cell target or as random introduction into a number of different cell types. However, the subject compositions will only have an effect in those cells in which the

target dsDNA is being transcribed or there is some other mechanism whereby the binding of the subject compositions can affect the mechanism. In this way selectivity can be achieved, since the only productive result will be in cells where the target dsDNA has an effect which is modified by the binding of the subject compositions to the dsDNA.

5

The subject compounds may be prepared, conveniently employing a solid support. See, for example, Baird and Dervan, J. Am. Chem. Soc. 1996, 118, 6141; PCT application, US97/03332. For solid phase synthesis, the oligomer is grown on the solid phase attached to the solid phase by a linkage which can be cleaved by a single step  
10 process. The addition of an aliphatic amino acid at the C-terminus of the oligomers allows the use of Boc- $\beta$ -alanine-Pam-resin which is commercially available in appropriate substitution levels (0.2mmol/g). Aminolysis may be used for cleaving the polyamide from the support. In the case of the N-methyl 4-amino-2-carboxypyrrole and the N-methyl 4-amino-2-carboxyimidazole, the activated esters such as N-hydroxysuccinidyl, 1,2,3-  
15 hydroxybenzotriazolyl, or the like may be employed, with the amino groups protected by Boc or Fmoc, with the monomers added sequentially in accordance with conventional techniques. For further details, see the references cited in the related literature, which are incorporated herein by reference, as well as the Experimental section.

20 The following examples are offered by way of illustration, and not by way of limitation.

### EXPERIMENTAL

25 Example 1: Solid phase synthesis of polyamides containing imidazole and pyrrole amino acids<sup>1</sup>

Boc- $\beta$ -alanine-(4-carboxamidomethyl)-benzyl-ester-copoly(styrene-divinylbenzene) resin (Boc- $\beta$ -alanine-Pam-Resin),  
30 dicyclohexylcarbodiimide (DCC), hydroxybenzotriazole (HOBt), 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), Boc-glycine, and Boc- $\beta$ -alanine were purchased from Peptides International. N,N-diisopropylethylamine (DIEA), N,N-dimethylformamide (DMF),

---

<sup>1</sup>Baird & Dervan, J. Am. Chem. Soc., 1996, 118, 6141.

N-methylpyrrolidone (NMP), and DMSO/NMP were purchased from Applied Biosystems. Boc- $\gamma$ -aminobutyric acid was from NOVA Biochem, dichloromethane (DCM) and triethylamine (TEA) was reagent grade from EM, thiophenol (PhSH), dimethylaminopropylamine, trichloroacetyl chloride, N-methylpyrrole, and

5 N-methylimidazole from Aldrich, and trifluoroacetic acid (TFA) from Halocarbon. All reagents were used without further purification.

#### Monomer Syntheses

##### 10 4-Nitro-2-trichloroacetyl-1-methylpyrrole

To a well stirred solution of trichloroacetyl chloride (1 kg, 5.5 mole) in 1.5 liter ethyl ether in a 12 liter flask was added dropwise over a period of 3 h a solution of N-methylpyrrole (0.45 kg, 5.5 mole) in 1.5 liter anhydrous ethyl ether. The reaction was

15 stirred for an additional 3 hours and quenched by the dropwise addition of a solution of 400 g potassium carbonate in 1.5 liters water. The layers were separated and the ether layer concentrated in vacuo to provide 2-(trichloroacetyl)pyrrole (1.2 kg, 5.1 mol) as a yellow crystalline solid sufficiently pure to be used without further purification. To a cooled (-40°C) solution of 2-(trichloroacetyl) pyrrole (1.2 kg, 5.1 mol) in acetic anhydride

20 (6 L) in a 12 L flask equipped with a mechanical stirrer was added 440 mL fuming nitric acid over a period of 1 hour while maintaining a temperature of (-40°C). The reaction was carefully allowed to warm to room temperature and stir an additional 4 h. The mixture was cooled to -30°C, and isopropyl alcohol (6 L) added. The solution was stirred at -20°C for 30 min during which time a white precipitate forms. The solution was

25 allowed to stand for 15 min and the resulting precipitate collected by vacuum filtration.

##### Methyl 4-nitropyrrole-2-carboxylate

To a solution of 4-Nitro-2-trichloroacetyl-1-methylpyrrole (800 g, 2.9 mol) in 2.5 L methanol in a 4 L Erlenmeyer flask equipped with a mechanical stirrer was added

30 dropwise a solution of NaH (60% dispersion in oil) (10g, 0.25 mol) in 500 mL methanol. The reaction was stirred 2 h at room temperature, and quenched by the addition of conc. sulfuric acid (25 mL). The reaction was then heated to reflux, allowed to slowly cool to room temperature. Product crystallized as white needles which were collected by vacuum

filtration.

35

Methyl 4-amino-1-methyl-pyrrole-2-carboxylate hydrochloride

Methyl 4-nitropyrrole-2-carboxylate 4 (450 g, 2.8 mol) was dissolved in ethyl acetate (8 L). A slurry of 40 g of 10% Pd/C in 800 mL ethyl acetate was then added and the mixture stirred under a slight positive pressure of hydrogen (c.a. 1.1 atm) for 48 h.

- 5 Pd/C was removed by filtration through celite, washed 1 x 50 mL ethyl acetate, and the volume of the mixture reduced to c.a. 500 mL. 7 L of cold ethyl ether was added and HCl gas gently bubbled through the mixture. The precipitated amine hydrochloride was then collected by vacuum filtration to yield a white powder (380 g, 81.6 %).

10 4-[(tert-Butoxycarbonyl)amino]-1-methylpyrrole-2-carboxylic acid

Methyl 4-amino-1-methyl-pyrrole-2-carboxylate hydrochloride (340 g, 1.8 mol) was dissolved in 1 L of 10% aqueous sodium carbonate in a 3 L flask equipped with a mechanical stirrer, di-t-butylidicarbonate (400 g, 2.0 mmol) slurried in 500 mL of dioxane was added over a period of thirty min., maintaining a temperature of 20°C. The reaction  
15 was allowed to proceed for three h and was determined complete by TLC, cooled to 5°C for 2 h and the resulting white precipitate collected by vacuum filtration. The Boc-pyrrole ester contaminated with Boc-anhydride was dissolved in 700 mL MeOH, 700 mL of 2M NaOH was added and the solution heated at 60°C for 6 h. The reaction was cooled to room temperature, washed with ethyl ether (4 x 1000 mL), the pH of the aqueous layer  
20 reduced to c.a. 3 with 10% (v/v) H<sub>2</sub>SO<sub>4</sub>, and extracted with ethyl acetate (4 x 2000 mL). The combined ethyl acetate extracts were dried (sodium sulfate) and concentrated in vacuo to provide a tan foam. The foam was dissolved in 500 mL of DCM and 2 L petroleum ether added, the resulting slurry was concentrated in vacuo. The reaction was redissolved and concentrated three additional times to provide a fine white powder (320 g, 78 %  
25 yield).

1,2,3-Benzotriazol-1-yl 4-[(tert-butoxycarbonyl)-amino]-1-methylpyrrole-2-carboxylate

- Boc-Py-acid (31 g, 129 mmol) was dissolved in 500 mL DMF, HOBt (17.4 g, 129  
30 mmol) was added followed by DCC (34 g, 129 mmol). The reaction was stirred for 24 h and then filtered dropwise into a well stirred solution of 5 L of ice water. The precipitate was allowed to sit for 15 min at 0°C and then collected by filtration. The wet cake was dissolved in 500 mL DCM, and the organic layer added slowly to a stirred solution of cold petroleum ether (4°C). The mixture was allowed to stand at -20°C for 4 h and then

collected by vacuum filtration and dried in vacuo to provide a finely divided white powder (39 g, 85% yield).

Ethyl 1-methylimidazole-2-carboxylate

- 5 N-methylimidazole (320 g, 3.9 mol) was combined with 2 L acetonitrile and 1 L triethylamine in a 12 L flask equipped with a mechanical stirrer and the solution cooled to -20°C. Ethyl chloroformate (1000 g, 9.2 mol) was added with stirring, keeping the temperature between -20°C and -25°C. The reaction was allowed to slowly warm to room temperature and stir for 36 h. Precipitated triethylamine hydrochloride was removed  
10 by filtration and the solution concentrated in vacuo at 65°C. The resulting oil was purified by distillation under reduced pressure (2 torr, 102°C) to provide a white solid (360 g, 82% yield).

Ethyl 1-methyl-4-nitroimidazole-2-carboxylate

- 15 Ethyl 1-methylimidazole-2-carboxylate was carefully dissolved in 1000 mL of concentrated sulfuric acid cooled to 0°C. 90% nitric acid (1 L) was slowly added maintaining a temperature of 0°C. The reaction was then refluxed with an efficient condenser (-20°C) in a well ventilated hood for 50 min. The reaction was cooled with an ice bath, and quenched by pouring onto 10 L ice. The resulting blue solution was then  
20 extracted with 20 L DCM, the combined extracts dried (sodium sulfate) and concentrated in vacuo to yield a tan solid which was recrystallized from 22 L of 21:1 carbon tetrachloride/ethanol. The resulting white crystals are collected by vacuum filtration.

Ethyl 4-amino-1-methylimidazole-2-carboxylate hydrochloride

- 25 Ethyl 1-methyl-4-nitroimidazole-2-carboxylate (103 g, 520 mmol) was dissolved in 5 L of 1:1 ethanol/ethyl acetate. 20 g 10% Pd/C slurried in 500 mL ethyl acetate was added and the mixture stirred under a slight positive pressure of hydrogen (c.a. 1.1 atm) for 48 h. The reaction mixture was filtered, concentrated in vacuo to a volume of 500 mL and 5 L of cold anhydrous ethyl ether added. Addition of HCl gas provided a white  
30 precipitate. The solution was cooled at -20°C for 4 h and the precipitate collected by vacuum filtration and dried in vacuo to provide (75 g, 78% yield) of a fine white powder.

4-[(tert-butoxycarbonyl)amino]-1-methylimidazole-2-carboxylic acid

- Ethyl 4-amino-1-methylimidazole-2-carboxylate hydrochloride (75 g, 395 mmol) was dissolved in 200 mL DMF. DIEA (45 mL, 491 mmol) was added followed by di-t-butylidicarbonate (99 g, 491 mmol). The mixture was shaken at 60°C for 18 h, allowed to assume room temperature, and partitioned between 500 mL brine, 500 mL ethyl ether. The ether layer was extracted with (2 x 200 mL each) 10% citric acid, brine, satd. sodium bicarbonate and brine, dried over sodium sulfate and concentrated in vacuo to yield the Boc-ester contaminated with 20% Boc-anhydride as indicated by <sup>1</sup>H NMR. The Boc-ester, used without further purification, was dissolved in 200 mL 1 M NaOH. The reaction mixture was allowed to stand for 3 h at 60°C with occasional agitation. The reaction mixture was cooled to 0°C, and carefully neutralized with 1 M HCl to pH 2, at which time a white gel formed. The gel was collected by vacuum filtration, frozen before drying, and remaining water lyophilized to yield a white powder.

4-[(tert-butoxycarbonyl)amino]-1-methylpyrrole-2-(4carboxamide-methyl imidazole)-2-carboxylic acid

- This compound was prepared as described below for (-[(tert-butoxycarbonyl)amino]-butyric acid (4-carboxamido-1-methyl-imidazole)-2-carboxylic acid, substituting Boc-Pyrrole acid for Boc-γ-aminobutyric acid. (4.1 g, 91% yield).

(γ-[(tert-butoxycarbonyl)amino]-butyric acid-(4-carboxamido-1-methyl-imidazole)-2-carboxylic acid

- To a solution of Boc-(γ-aminobutyric acid (10 g, 49 mmol) in 40 mL DMF was added 1.2 eq HOBt (7.9 g, 59 mmol) followed by 1.2 eq DCC (11.9 g, 59 mmol). The solution was stirred for 24 h, and the DCU removed by filtration. Separately, to a solution of ethyl 4-nitro-1-methylimidazole-2-carboxylate (9.8 g, 49 mmol) in 20 mL DMF was added Pd/C catalyst (10%, 1 g), and the mixture was hydrogenated in a Parr bomb apparatus (500 psi H<sub>2</sub>) for 2 h. The catalyst was removed by filtration through celite and filtrate immediately added to the HOBt ester solution. An excess of DIEA (15 mL) was then added and the reaction stirred at 37°C for 48 h. The reaction mixture was then added dropwise to a stirred solution of ice water and the resulting precipitate collected by vacuum filtration to provide crude ethyl 4-[[[3-[(tert-butoxycarbonyl)amino]propyl] carbonylamino]-1-methylimidazole-2-carboxylate (5 g, 14.1



mmol). To the crude ester dissolved in 50 mL methanol was added 50 mL 1M KOH and the resulting mixture allowed to stir for 6 h at 37°C. Excess methanol was removed in vacuo and the resulting solution acidified by the addition of 1 M HCl. The resulting precipitate was collected by vacuum filtration and dried in vacuo to yield a brown powder.  
5 (4.4g, 89% yield).

#### Solid Phase Syntheses

Activation of Imidazole-2-carboxylic acid, ( $\gamma$ -aminobutyric acid, Boc-glycine, and Boc- $\beta$ -alanine. The appropriate amino acid or acid (2 mmol) was dissolved in 2 mL  
10 DMF. HBTU (720 mg, 1.9 mmol) was added followed by DIEA (1 mL) and the solution lightly shaken for at least 5 min.

#### Activation of Boc-Imidazole acid

Boc-imidazole acid (257 mg, 1 mmol) and HOBt (135 mg, 1 mmol) were  
15 dissolved in 2 mL DMF, DCC (202 mg, 1 mmol) is then added and the solution allowed to stand for at least 5 min.

#### Activation of Boc- $\gamma$ -Imidazole acid and Boc-Pyrrole-Imidazole acid

The appropriate dimer (1 mmol) and HBTU (378 mg, 1 mmol) are combined in 2  
20 mL DMF. DIEA (1 mL) is then added and the reaction mixture allowed to stand for 5 min.

#### Activation of Boc-Pyrrole acid (for coupling to Imidazole amine)

Boc-Pyrrole acid (514 mg, 2 mmol) was dissolved in 2 mL dichloromethane, DCC  
25 (420 mg, 2 mmol) added, and the solution allowed to stand for 10 min, DMAP (101 mg, 1 mmol) was added and the solution allowed to stand for 1 min.

#### Acetylation Mix

2 mL DMF, DIEA (710  $\mu$ L, 4.0 mmol), and acetic anhydride (380  $\mu$ L, 4.0 mmol)  
30 were combined immediately before use.

#### Manual Synthesis Protocol

Boc- $\beta$ -alanine-Pam-Resin (1.25 g, 0.25 mmol) is placed in a 20 mL glass reaction vessel, shaken in DMF for 5 min and the reaction vessel drained. The resin was washed

with DCM (2 x 30 s.) and the Boc group removed with 80% TFA/DCM/0.5 M PhSH, 1 x 30 s., 1 x 20 min. The resin was washed with DCM (2 x 30 s.) followed by DMF (1 x 30 s.). A resin sample (5 - 10 mg) was taken for analysis. The vessel was drained completely and activated monomer added, followed by DIEA if necessary. The reaction  
5 vessel was shaken vigorously to make a slurry. The coupling was allowed to proceed for 45 min, and a resin sample taken. The reaction vessel was then washed with DCM, followed by DMF.

#### Machine-Assisted Protocols

10 Machine-assisted synthesis was performed on a ABI 430A synthesizer on a 0.18 mmol scale (900 mg resin; 0.2 mmol/gram). Each cycle of amino acid addition involved: deprotection with approximately 80% TFA/DCM/0.4 M PhSH for 3 minutes, draining the reaction vessel, and then deprotection for 17 minutes; 2 dichloromethane flow washes; an  
15 NMP flow wash; draining the reaction vessel; coupling for 1 hour with in situ neutralization, addition of dimethyl sulfoxide (DMSO)/NMP, coupling for 30 minutes, addition of DIEA, coupling for 30 minutes; draining the reaction vessel; washing with DCM, taking a resin sample for evaluation of the progress of the synthesis by HPLC analysis; capping with acetic anhydride/DIEA in DCM for 6 minutes; and washing with  
20 DCM. A double couple cycle is employed when coupling aliphatic amino acids to imidazole, all other couplings are performed with single couple cycles.

The ABI 430A synthesizer was left in the standard hardware configuration for NMP-HOBt protocols. Reagent positions 1 and 7 were DIEA, reagent position 2 was TFA/0.5 M thiophenol, reagent position 3 was 70% ethanolamine/methanol, reagent  
25 position 4 was acetic anhydride, reagent position 5 was DMSO/NMP, reagent position 6 was methanol, and reagent position 8 was DMF. New activator functions were written, one for direct transfer of the cartridge contents to the concentrator (switch list 21, 25, 26, 35, 37, 44), and a second for transfer of reagent position 8 directly to the cartridge (switch list 37, 39, 45, 46).

30

Boc-Py-OBt ester (357 mg, 1 mmol) was dissolved in 2 mL DMF and filtered into a synthesis cartridge. Boc-Im acid monomer was activated (DCC/HOBt), filtered, and placed in a synthesis cartridge. Imidazole-2-carboxylic acid was added manually. At the initiation of the coupling cycle the synthesis was interrupted, the reaction vessel vented

and the activated monomer added directly to the reaction vessel through the resin sampling loop via syringe. When manual addition was necessary an empty synthesis cartridge was used. Aliphatic amino acids (2 mmol) and HBTU (1.9 mmol) were placed in a synthesis cartridge. 3 mL of DMF was added using a calibrated delivery loop from reagent bottle  
5 8, followed by calibrated delivery of 1 mL DIEA from reagent bottle 7, and a 3 minute mixing of the cartridge.

The activator cycle was written to transfer activated monomer directly from the cartridge to the concentrator vessel, bypassing the activator vessel. After transfer, 1 mL  
10 of DIEA was measured into the cartridge using a calibrated delivery loop, and the DIEA solution combined with the activated monomer solution in the concentrator vessel. The activated ester in 2:1 DMF/DIEA was then transferred to the reaction vessel. All lines were emptied with argon before and after solution transfers.

15 ImPyPy- $\gamma$ -PyPyPy- $\beta$ -alanine-Dp

ImPyPy- $\gamma$ -PyPyPy- $\beta$ -alanine-Pam-Resin was prepared by machine-assisted synthesis protocols. A sample of resin (1 g, 0.17 mmol) was placed in a 20 mL glass scintillation vial, 4 mL of dimethylaminopropylamine added, and the solution heated at 55°C for 18 h. Resin is removed by filtration through a disposable propylene filter and  
20 mL of water added. The polyamide/amine mixture was purified directly by preparatory HPLC and the appropriate fractions lyophilized to yield a white powder.

Stepwise HPLC analysis

A resin sample (c.a. 4 mg) was placed in a 4 mL glass test tube, 200  $\mu$ L of  
25 N,N-dimethylaminopropylamine was added and the mixture heated at 100°C for 5 min. The cleavage mixture was filtered and a 25  $\mu$ L sample analyzed by analytical HPLC at 254 nm.

Example 2:  $\beta$ -alanine and  $\gamma$ -aminobutyric acid are turn and overlapped specific "guide" amino acids which may be combined predictably within the same molecule.<sup>2</sup>

Synthesis of Polyamides ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-Dp and

5 ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp

- All polyamides were prepared in high purity using solid phase synthetic methodology as described above. Polyamides were assembled in a stepwise manner on Boc- $\beta$ -alanine-Pam resin and Boc-glycine-Pam-resin respectively. Polyamides ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-Dp, ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-  
 10 PyPyPy-G-Dp and ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp-NH<sub>2</sub> were cleaved from the support with an appropriate primary amine and purified by reversed-phase HPLC to provide 10-30 mg of polyamide. A primary amine group suitable for post-synthetic modification can be provided. Amine modified polyamides are treated with an excess of the dianhydride of EDTA, unreacted anhydride hydrolyzed, and the  
 15 EDTA modified polyamide ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp-EDTA isolated by reversed-phase HPLC.

ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp-NH<sub>2</sub>

- Polyamide was prepared by machine-assisted solid phase methods as a white  
 20 powder (29 mg, 59% recovery).

ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp-EDTA

- EDTA-dianhydride (50 mg) was dissolved in 1 mL DMSO/NMP solution and 1 mL DIEA by heating at 55°C for 5 min. The dianhydride solution was added to  
 25 ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp-NH<sub>2</sub> (9.0 mg, 5  $\mu$ mol) dissolved in 750  $\mu$ L DMSO. The mixture was heated at 55°C for 25 min, treated with 3 mL 0.1 M NaOH, and heated at 55°C for 10 min. 0.1% TFA was added to adjust the total volume to 8 mL and the solution purified directly by reversed-phase HPLC to provide ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp-EDTA as a white powder  
 30 (3 mg, 30% recovery after HPLC purification).

---

<sup>2</sup>Tranger et al, Chemistry and Biology, 1996, 3, 369-377.

Preparation of <sup>32</sup>P-labeled DNA

Plasmid pJT8 was prepared by hybridizing two sets of 5'-phosphorylated complementary oligonucleotides,  
5'-CCGGGAACGTAGCGTACCGGTCGCAAAAAGACAGGCTCGA-3', and  
5'-GGCGTCGAGCCTGTCTTTTGGCGACCGGTACGCTACGTTC-3', and  
5'-CGCCGCATATAGACAGGCCAGCTGCGTCCTAGCTAGCGTCGTAGCGTCTTA  
AGAG-3' and  
5'-TCGACTCTTAAGACGCTACGACGCTAGCTAGGACGCAGCTGGGCCTGTCTAT  
ATGC-3', and ligating the resulting duplexes to the large pUC19 *Ava*I/*Sal*I restriction  
fragment. The 3'-<sup>32</sup>P end-labeled *Afl*III/*Fsp*I fragment was prepared by digesting the  
plasmid with *Afl*III and simultaneously filling in using Sequenase,  
[" $\alpha$ -<sup>32</sup>P]-deoxyadenosine-5'-triphosphate, and [" $\alpha$ -<sup>32</sup>P]-thymidine-5'-triphosphate, digesting  
with *Fsp*I, and isolating the 247 bp fragment by nondenaturing gel electrophoresis. The  
5'-<sup>32</sup>P-end-labeled *Afl*III/*Fsp*I fragment was prepared using standard methods. A and G  
sequencing were carried out as described.<sup>3</sup> Standard methods were used for all DNA  
manipulations.<sup>4</sup>

Affinity cleavage reactions

All reactions were executed in a total volume of 400  $\mu$ L. A stock solution of  
EDTA-modified polyamide or H<sub>2</sub>O was added to a solution containing labeled restriction  
fragment (15,000 cpm), affording final solution conditions of 20 mM HEPES, 200 mM  
NaCl, 50  $\mu$ g/mL glycogen, and pH 7.3. Subsequently, 20  $\mu$ L of freshly prepared 20 mM  
Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub> was added and the solution allowed to equilibrate for 20 min. Cleavage  
reactions were initiated by the addition of 40  $\mu$ L of 50 mM dithiothreitol, allowed to  
proceed for 12 min at 22°C, then stopped by the addition of 1 mL of ethanol. Reactions  
were precipitated and the cleavage products separated using standard methods. Next, 10  
 $\mu$ L of a solution containing calf thymus DNA (140  $\mu$ M base-pair) (Pharmacia) and  
glycogen (2.8 mg/mL) was added, and the DNA precipitated. The reactions were  
resuspended in 1x TBE/80% formamide loading buffer, denatured by heating at 85°C for

---

<sup>3</sup>Maxam & Gilbert, *Methods Enzymol.*, 1980, 65, 499-560; Iverson & Dervan, *Methods Enzymol.*, 1987, 15, 7823-7830; Sambrook et al, 1989, *Molecular Cloning*, 2nd ed., Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

<sup>4</sup> Sambrook et al, 1989, *Molecular Cloning*, 2nd ed., Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY.

10 min, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% crosslink, 7 M urea) in 1x TBE at 2000 V. Gels were dried and exposed to a storage phosphor screen. Relative cleavage intensities were determined by volume integration of individual cleavage bands using ImageQuant software.

5

#### Quantitative DNase I footprint titration experiments

All reactions were executed in a total volume of 400  $\mu$ L. A polyamide stock solution or H<sub>2</sub>O (for reference lanes) was added to an assay buffer containing radiolabeled restriction fragment (15,000 cpm), affording final solution conditions of 10 mM TrisHCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, pH 7.0, and either (i) 1 pM - 10 nM polyamide or (ii) no polyamide (for reference lanes). The solutions were allowed to equilibrate at 22°C for (i) 12 h for polyamide 1 or (ii) 36 h for polyamide 2. Footprinting reactions were initiated by the addition of 10  $\mu$ L of a DNase I stock solution (at the appropriate concentration to give ~ 55% intact DNA) containing 1 mM dithiothreitol and allowed to proceed for seven min at 22°C. The reactions were stopped by the addition of 50  $\mu$ L of a solution containing 2.25 M NaCl, 150 mM EDTA, 0.6 mg/mL glycogen, and 30  $\mu$ M base-pair calf thymus DNA, and ethanol precipitated. Reactions were resuspended in 1x TBE/80% formamide loading buffer, denatured by heating at 85°C for 10 min, and placed on ice. The reaction products were separated by electrophoresis on an 8% polyacrylamide gel (5% crosslink, 7 M urea) in 1x TBE at 2000 V. Gels were dried and exposed to a storage phosphor screen (Molecular Dynamics).

#### Quantitation and data analysis

Data from the footprint titration gels were obtained using a Molecular Dynamics 400S PhosphorImager followed by quantitation using ImageQuant software (Molecular Dynamics). Background-corrected volume integration of rectangles encompassing the footprint sites and a reference site at which DNase I reactivity was invariant across the titration generated values for the site intensities ( $I_{site}$ ) and the reference intensity ( $I_{ref}$ ). The apparent fractional occupancy ( $Q_{app}$ ) of the sites were calculated using the equation (1):

30

$$Q_{app} = 1 - \frac{I_{site}/I_{ref}}{I_{site}^{\circ}/I_{ref}^{\circ}} \quad (1)$$

where  $I_{o_{site}}$  and  $I_{o_{ref}}$  are the site and reference intensities, respectively, from a control lane to which no polyamide was added. The  $([L]_{tot}, Q_{app})$  data points were fit to a general Hill equation (eq 2) by minimizing the difference between  $Q_{app}$  and  $Q_{fit}$ :

$$Q_{fit} = Q_{min} + (Q_{max} - Q_{min}) \frac{K_a^n [L]_{tot}^n}{1 + K_a^n [L]_{tot}^n} \quad (2)$$

- 5 where  $[L]_{tot}$  is the total polyamide concentration,  $K_a$  is the equilibrium association constant, and  $Q_{min}$  and  $Q_{max}$  are the experimentally determined site saturation values when the site is unoccupied or saturated, respectively. The data were fit using a nonlinear least-squares fitting procedure with  $K_a$ ,  $Q_{max}$ , and  $Q_{min}$  as the adjustable parameters. For polyamide ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-Dp, binding isotherms for the
- 10 5'-AGACA-3' target sites were adequately fit by Langmuir isotherms (eq 2,  $n=1$ ), consistent with formation of 1:1 polyamide-DNA complexes. For ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp, steeper binding isotherms (eq 2,  $n=1.8-2.2$ ) were observed at the target sites 5'-AAAAAGACA-3' and 5'-ATATAGACA-3'. The steepness of these isotherms may be due to the very high equilibrium association constants at these
- 15 sites. Treatment of the data in this manner does not represent an attempt to model a binding mechanism. The data is a comparison of values of the apparent first-order association constant, a value that represents the concentration of ligand at which a site is half-saturated. The binding isotherms were normalized using the following equation:

$$Q_{norm} = \frac{Q_{app} - Q_{min}}{Q_{max} - Q_{min}} \quad (3)$$

- 20 Four sets of data were used in determining each association constant. The method for determining association constants used here involves the assumption that  $[L]_{tot} = [L]_{free}$ , where  $[L]_{free}$  is the concentration of polyamide free in solution (unbound). For very high association constants this assumption becomes invalid, resulting in underestimated association constants. In the experiments described here, the DNA concentration is
- 25 estimated to be  $\sim 5$  pM. As a consequence, apparent association constants greater than  $\sim 10^{10} M^{-1}$  should be regarded as lower limits.

DNA-binding orientation

Affinity cleavage<sup>5</sup> experiments with ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp-EDTAFE(II) (2-Fe(II)) on the 5'- or 3'-<sup>32</sup>P end-labeled 247 bp pJT4 AflIII/FspI restriction fragment revealed that this polyamide selectively binds the

5 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' target sequences at subnanomolar concentration. A single 3'-shifted cleavage pattern is observed at each 9 bp site indicating that the polyamide is bound in one orientation with the C-terminus at the 5' end of the 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' sequences.

10 DNA-binding affinity and specificity

The exact locations and sizes of all binding sites were determined first by preliminary MPEFe(II) footprinting experiments.<sup>6</sup> Quantitative DNase I footprint titration experiments<sup>7</sup> on the 3'-<sup>32</sup>P-labeled 247 bp restriction fragment (10 mM TrisHCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, pH 7.0, 22°C) reveal that ImPyPy- $\gamma$ -aminobutyric

15 acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp specifically binds 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' with equilibrium association constants of  $K_a = 2 \times 10^{10} \text{ M}^{-1}$  and  $K_a = 8 \times 10^9 \text{ M}^{-1}$ , respectively. Additional sites on the restriction fragment are bound with lower affinity. For comparison, the six-ring hairpin polyamide ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-Dp binds 5'-aaaaAGACA-3' and 5'-atatAGACA-3' with

20 association constants of  $K_a = 5 \times 10^7 \text{ M}^{-1}$  and  $K_a = 9 \times 10^7 \text{ M}^{-1}$ , respectively.

Relative to the six-ring polyamide ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-Dp, the nine-ring polyamide ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp binds 5'-AAAAAGACA-3' and 5'-ATATAGACA-3' with

25 ~400-fold and ~100-fold higher affinity, respectively. Similar binding enhancements have recently been reported in a separate system.<sup>8</sup> Addition of a C-terminal PyPyPy subunit

---

<sup>5</sup>Wade et al, J. Am. Chem. Soc., 1992 114, 8783-8794; Schultz et al, J. Am. Chem. Soc., 1982, 104, 6861-6863.

<sup>6</sup>Hertzberg & Dervan, J. Am. Chem. Soc., 1982 104, 313-315.

<sup>7</sup>Fox & Waring, Nucleic Acids Res., 1984, 12, 9271-9285; Brenowitz et al, Methods Enzymol., 1986, 130, 132-181; Brenowitz, et al, Proc. Natl. Acad. Sci. U.S.A., 1986, 83, 8462-8466.

<sup>8</sup>Trauger et al., Nature, 1996, 382, 559-561



using a  $\beta$ -alanine linker is an effective strategy for increasing the DNA-binding affinity of hairpin polyamides that bind adjacent to an (A,T)<sub>4</sub> sequence.

Polyamide ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-PyPyPy-G-Dp binds  
5 several mismatch sites present on the 247 bp restriction fragment with high affinity. The two highest affinity mismatch sites, 5'-GAATTCAC-3' ( $K_a = 4.5 \times 10^9 \text{ M}^{-1}$ ) and 5'-GTTTTCCCA-3' ( $K_a = 2.5 \times 10^9 \text{ M}^{-1}$ ), are bound with at least 5-fold reduced affinity relative to the optimal match site 5'-AAAAAGACA-3' (formally mismatched base-pairs are highlighted), although this value may be a lower limit due to the uncertainty in the  
10 very high equilibrium association constant for the optimal match site. In contrast, the six-ring polyamide ImPyPy- $\gamma$ -aminobutyric acid-ImPyPy- $\beta$ -alanine-Dp binds more strongly to the match site 5'-AGACA-3' over the single base-pair mismatch sites 5'-ATTCA-3' and 5'-TTACA-3' by a factor of 10.

15 Example 3. Subnanomolar ligand binding with single mismatch differentiation

The DNA-binding affinities were evaluated of two eight-ring hairpin polyamides, ImPyPyPy- $\gamma$ -ImPyPyPy- $\beta$ -Dp (1) and ImPyPyPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp (2), which differ by a single amino acid, for two 6 base pair (bp) target sites, 5'-AGTACT-3' and 5'-AGTATT-3', which differ by a single base pair. Based on the pairing rules for  
20 polyamide-DNA complexes, the sites 5'-AGTACA-3' and 5'-AGTATT-3' are for polyamide 1 "match" and "single base pair mismatch" sites, respectively, and for polyamide 2 "single base pair mismatch" and "match" sites, respectively (Figure 2).

Polyamides 1 and 2 were synthesized by solid phase methods and purified by  
25 reversed phase HPLC<sup>9</sup>. The identity and purity of the polyamides was verified by <sup>1</sup>H NMR, MALDI-TOF MS, and analytical HPLC. MALDI-TOF MS: 1, 1223.4 (1223.3 calculated for M+H); 2, 1222.3 (1222.3 calculated for M+H). Equilibrium association constants for complexes of 1 and 2 with match and mismatch six base pair binding sites on a 3'-<sup>32</sup>P-labeled 229 bp restriction fragment were determined by  
30 quantitative DNase I footprint titration experiments<sup>10</sup> (Table I).

---

<sup>9</sup>Baird and Dervan, J. Am. Chem. Soc., 1996, 118, 6141-6146

<sup>10</sup>Galas & Schmitz, Nucleic Acids Res., 1978, 5 3157-3170; Fox & Waring, ibid, 1984, 12, 9271-9285; Brenowitz et al., Meth. Enzym. 1986, 130, 132-181

TABLE 1 Equilibrium association constants ( $M^{-1}$ )

Binding Site	1	2
5'-ttAGTACTtg-3'	$3.7 \times 10^{10}$ (0.8)	$5.0 \times 10^8$ (0.5)
5'-ttAGTATTtg-3'	$4.1 \times 10^8$ (0.5)	$3.5 \times 10^9$ (0.8)

10 The reported association constants are the average values obtained from three DNase I footprint titration experiments. The standard deviation for each data set is indicated in parentheses. Assays were carried out in the presence of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, and 5 mM CaCl<sub>2</sub> at pH 7.0 and 22 °C. The six base-pair binding sites are in capital letters, with flanking sequences in lower-case letters.

15 Polyamide 1 binds its match site 5'-AGTACT-3' at 0.03 nM concentration and its single base pair mismatch site 5'-AGTATT-3' with nearly 100-fold lower affinity. Polyamide 2 binds its designated match site 5'-AGTATT-3' at 0.3 nM concentration and its single base pair mismatch site 5'-AGTACT-3' with nearly 10-fold lower affinity. The specificity of 1 and 2 for their respective match sites results from very small structural  
20 changes. Replacing a single nitrogen atom in 1 with C-H (as in 2) reduces the affinity of the polyamide•5'-AGTACT-3' complex by ~75-fold representing a free energy difference of ~2.5 kcal/mole. Similarly, replacing a C-H in 2 with N (as in 1) reduces the affinity of the polyamide•5'-AGTATT-3' complex ~10-fold, a loss in binding energy of ~1.3 kcal/mol.

25 Quantitative DNase I footprint titration experiments with polyamides 1 and 2 on the 3'-<sup>32</sup>P-labeled 229 bp pJT8 AflIII/FspI restriction fragment. Comparison lanes were A and G sequencing lanes; DNase I digestion products obtained in the absence of polyamide; DNase I digestion products obtained in the presence of 1 pM, 2 pM, 5 pM, 10 pM, 15  
30 pM, 25 pM, 40 pM, 65 pM, 0.1 nM, 0.15 nM, 0.25 nM, 0.4 nM, 0.65 nM, 1 nM, 2 nM, 5 nM, and 10 nM polyamide, respectively; and intact DNA. Polyamide binding sites for which association constants were determined for 5'-AGTACT-3' and 5'-AGTATT-3'. Additional sites not analyzed were 5'-TGTAAG-3', 5'-TGTGCT-3', and 5'-TAAGTT-3'. All reactions were executed in a total volume of 400 μL. A polyamide stock solution or  
35 H<sub>2</sub>O was added to an assay buffer containing radiolabeled restriction fragment, affording final solution conditions of 10 mM Tris•HCl, 10 mM KCl, 10 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub>, and pH 7.0. The solutions were allowed to equilibrate for 12-15 h at 22 °C prior to initiation of footprinting reactions. Footprinting reactions, separation of cleavage

- products, and data analysis were carried out as described elsewhere<sup>11</sup>. Plasmid pJT8 was prepared by hybridizing two 5'-phosphorylated complementary oligonucleotides, 5'-CCGGTTAGTATTTGGATGGGCCTGGTTAGTA-CTTGGATGGGAGACCGCCTGGGAATACCAGGTGTCGTATCTTAAGAG-3' and 5'-TCGACTCTTAAGATACGACACCTGGTATTCCCAGGCGGTCTCCCATCCAA-GTACTAACCAGGCCCATCCAAATACTAA-3', and ligating the resulting duplex to the large pUC19 *Ava*I/*Sal*I restriction fragment.

#### Example 4. Intracellular binding and transcription inhibition

##### 10 Methods

- Polyamides.** Polyamides were synthesized by solid phase methods.<sup>12</sup> The identity and purity of the polyamides was verified by <sup>1</sup>H NMR, matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS), and analytical HPLC. MALDI-TOF-MS: 1, 1223.4 (1223.3 calcd for M+H); 2, 1222.3 (1222.3 calcd for M+H); 3, 1223.1 (1223.3 calcd for M+H).

- Transcription inhibition in vitro. A high speed cytosolic extract from unfertilized *Xenopus* eggs was prepared as described.<sup>13</sup> DNA templates for transcription were the somatic-type 5S RNA gene contained in plasmid pX1s11<sup>14</sup> (50 ng per reaction) and the *tyrD* tRNA gene contained in plasmid pTyrD<sup>15</sup> (100 ng plasmid DNA per reaction), both from *X. laevis*.
- 20 Transcription reactions (20  $\mu$ L final volume) contained the following components: 2.5  $\mu$ L extract, 9 ng (12 nM) of TFI<sub>II</sub>A isolated from immature oocytes<sup>16</sup>, 0.6 mM ATP, UTP, CTP, 0.02 mM GTP and 10  $\mu$ Ci of [ $\alpha$ -<sup>32</sup>P] GTP and the final buffer components 12 mM HEPES (pH 7.5), 60 mM KCl, 6 mM MgCl<sub>2</sub>, 25  $\mu$ M ZnCl<sub>2</sub>, and 8% (v/v) glycerol. Plasmid DNAs were pre-incubated with polyamides in the same buffer prior to adding
- 25 TFI<sub>II</sub>A and other reaction components. RNA was purified and analyzed on a denaturing 6% polyacrylamide gel. A Molecular Dynamics Phosphorimager equipped with ImageQuant software was used to quantify the effect of the polyamides on 5S and tRNA gene transcription.

<sup>11</sup>Mrkisch et al., *J. Am. Chem. Soc.*, 1994, 116, 7983-7988

<sup>12</sup>Baird, E. E. and Dervan, P. B., *J. Am. Chem. Soc.* 118, 6141-6146 (1996)

<sup>13</sup>Hartl, P. et al., *J. Cell Biol.* 120, 613-624 (1993)

<sup>14</sup>Peterson, R. C. et al., *Cell* 20, 131-144 (1980)

<sup>15</sup>Stutz, F. et al., *Genes Dev.* 3, 1190-1198 (1989)

<sup>16</sup>Smith, D. R. et al., *Cell* 37, 645-652 (1984)

**Transcription inhibition in vivo.** Fibroblasts from a *Xenopus* kidney derived cell line (kindly provided by Dr. P. Labhart, Scripps) were grown at ambient temperature in 25 cm<sup>2</sup> culture flasks in Dulbecco's modified Eagle medium containing 10% (v/v) fetal calf serum. Cells were passaged for a minimum of three days prior to the addition of

5 polyamide to the culture medium. Incubations were continued for various times and nuclei were prepared by hypotonic lysis and used as templates for transcription as described.<sup>17</sup> DNA content was determined by measuring the absorbance of an aliquot of the isolated nuclei in 1% (w/v) sodium dodecyl sulfate (using an extinction coefficient at 260 nm of 1 AU = 50 µg/mL DNA). The buffer components and labeled and unlabeled nucleoside

10 triphosphates were as for the plasmid transcription reactions. Reactions were supplemented with 2 µL of RNA polymerase III (at approximately 50 µg/mL) isolated from *Xenopus* oocytes.<sup>18</sup>

### Results

15 The effect of polyamide 1 (ImPyPyPy-γ-ImPyPyPy-β-Dp) on TFI<sub>IIA</sub> binding to a restriction fragment isolated from a 5S RNA gene-containing plasmid was examined. Zf1-3, a recombinant TFI<sub>IIA</sub> analog missing fingers 4-9, binds in the major groove of the C-block promoter element (see Fig. 1). DNase I footprinting demonstrates that zfi-3 and polyamide 1 can co-occupy the same DNA molecule. When 5 nM polyamide 1 was

20 preincubated with the same DNA target, the binding of nine finger TFI<sub>IIA</sub> was inhibited by >90%. The differential inhibition of zfi-3 and full-length TFI<sub>IIA</sub> provides evidence that finger 4 interacts with or is placed in the minor groove. Polyamide 1 does not inhibit TFI<sub>IIA</sub> binding to 5S RNA.

25 Transcription of the 5S RNA gene in an in vitro system was monitored in the presence of increasing concentrations (10-60 nM) of polyamide 1. In these experiments, polyamide 1 was added to a 5S RNA gene containing plasmid prior to the addition of exogenous TFI<sub>IIA</sub> (12 nM) and a crude extract derived from unfertilized *Xenopus* eggs. As a control, a tyrosine tRNA gene was included on a separate plasmid in these reactions.

30 The tRNA gene has an upstream binding site for 1, but lacks a predicted protein-polyamide interaction. Both genes are actively transcribed in this system, either individually or in mixed template reactions. Addition of 60 nM polyamide 1 inhibits 5S

---

<sup>17</sup>Schlissel, M. S. and Brown, D. D., *Cell* 37, 903-913 (1984)

<sup>18</sup>Roeder, R. G., *J. Biol. Chem.* 258, 1932-1941 (1983)

gene transcription by > 80%. Only a small degree of non-specific inhibition of tRNA transcription is observed at the concentrations of polyamide 1 required for efficient 5S RNA inhibition. The targeted 5S RNA gene is inhibited approximately 10-fold more effectively than the control tRNA gene. Mismatch polyamides 2 (ImPyPyPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp) and 3 (ImPyImPy- $\gamma$ -PyPyPyPy- $\beta$ -Dp) do not inhibit 5S RNA transcription at concentrations up to 60 nM. If the TFIIIA-DNA complex is first allowed to form, 30 nM polyamide 1 added, and the mixture incubated for 90 minutes prior to adding egg extract, efficient inhibition (80%) of 5S RNA transcription is also observed. Shorter incubation times result in less inhibition. The required incubation time of 90 minutes is similar to the measured half-life of the TFIIIA-DNA complex and supports that polyamide 1 forms a more stable complex with DNA than does TFIIIA.

The effect of the polyamides on 5S gene transcription in vivo was monitored. *Xenopus* kidney-derived fibroblasts were grown in the presence of increasing concentrations of polyamide 1 in the culture medium for various times. We found that concentrations of polyamide up to 1  $\mu$ M were not toxic, as measured by cell density, if growth was limited to less than 72 hours. Nuclei were prepared from cells by hypotonic lysis and equivalent amounts of the isolated nuclei from control and treated cells were used as templates for transcription with exogenous RNA polymerase III and labeled and unlabeled nucleoside triphosphates. This experiment monitors the occupancy of class III genes with active transcription complexes.<sup>19</sup> 5S RNA transcription can easily be assessed since the repetitive 5S genes give rise to a prominent band on a denaturing polyacrylamide gel. An autoradiogram was taken of the gel and the following observations made based on the observed autoradiogram.

Concentrations of polyamide 1 as low as 100 nM have a pronounced and selective effect on 5S transcription. At higher polyamide concentration, a general decrease in the transcriptional activity of the nuclei is observed; however, at each concentration tested, the effects of the polyamide are much greater on 5S RNA transcription than on tRNA transcription. Having established that nearly maximal inhibition of 5S transcription is achieved with 1  $\mu$ M polyamide 1, we monitored nuclear transcription after various times of cell growth in the presence of the polyamide. No inhibition is observed for zero time

---

<sup>19</sup>Schlissel, M. S., *Cell*, *supra*

incubation with polyamide 1 at 1  $\mu$ M concentration, indicating that disruption of transcription complexes does not occur during or after the isolation or work-up of cell nuclei. Statistically equivalent levels of 5S transcription were observed when the cells were exposed to polyamide 1 for 24, 48 or 72 hours.

5

The observations support the conclusion that polyamide 1 is able to enter cells, transit to the nucleus and disrupt transcription complexes on the chromosomal 5S RNA genes. To rule out the possibility that the observed inhibition is due to some non-specific toxicity of the polyamide rather than to direct binding to the 5S RNA gene, the effects of mismatch polyamides 2 and 3 in the nuclear transcription assay were monitored. Only a small effect on 5S RNA synthesis relative to tRNA synthesis is observed with 1  $\mu$ M of the mismatch polyamides 2 or 3 in the culture medium for 24 hours. This result indicates that the general inhibition of transcription observed with high concentrations of polyamide 1 may be a secondary effect of the inhibition of 5S RNA synthesis in vivo, rather than the result of non-specific polyamide interactions. Polyamide 2 affects a small enhancement of 5S RNA transcription in vitro and in vivo, indicating that polyamides may be able to upregulate transcription in certain cases.

As evidenced by the above results, the subject invention provides novel compounds, which are oligomers of organic cyclic groups, particularly azoles, where the compounds fit in the minor groove of dsDNA and provide for hydrogen bonding, polar interactions, and van der Waal's interactions resulting in high affinities and high association constants.

The subject compositions provide for substantial differentiation between the target sequence and single mismatch sequences. Normally, there will be at least a two-fold difference between the two sequences, more usually at least a five-fold difference, and preferentially at least a ten-fold difference or greater. In this way, one can insure that the target sequence will be primarily affected, with little effect on other sequences. Normally, the target sequence will be at least five nucleotides, usually at least six nucleotides, more usually at least eight nucleotides, and not more than about twenty nucleotides. By using combinations of compositions, where the combinations bind to different sequences, which may be proximal to each other, one may further enhance the inhibition at a particular gene.

The subject compositions are shown to bind with high affinities to specific dsDNA sequences and with substantially lower affinities to single base mismatches. In this way, even in complex compositions of dsDNA, such as may be encountered in cellular compositions, there is substantial assurance that the target sequence will be affected and other sequences will be little affected, if at all. Furthermore, the subject compositions are capable of transport across a cellular membrane and through the cytosol to the nucleus. The subject compositions are capable of binding to chromosomal dsDNA involved with nucleosomes and inhibit transcription of genes which form complexes with the subject compositions. Single oligomers may be employed or combinations of oligomers to provide for the desired complex formation. By using the subject compositions in diagnosis, one is not required to melt the DNA to provide for single-stranded DNA. Rather, the subject compositions can accurately target the dsDNA and avoid the melting and competition between the natural strands and the labeled complementary strand, as is employed conventionally today. The subject compositions may be used for cleavage of dsDNA at specific sites, so as to isolate target DNA, which may then be readily amplified using PCR. By further modifying the subject compositions, one may further expand their applications in their use for identifying sequences, cleaving specific sequences, investigating the role of genes, screening for the presence of sequences in cells, and inhibiting proliferation of cells.

20

The references described throughout this specification are fully incorporated by reference.

Having now fully described the invention, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the invention as set forth herein.

25

**WHAT IS CLAIMED IS:**

1. A method for forming a specific complex between target dsDNA and from 1 to 2 oligomers at a minor groove site, where said oligomers are selected to provide binding to said target dsDNA at a  $K_a$  of at least  $10^9 M^{-1}$ ,  
said oligomers comprising as a core structure: (1) organic cyclic groups of from 5 to 6 annular members, where at least 60% of the rings are heterocycles having from 1 to 3 heteroatoms, where the heteroatoms are nitrogen, oxygen and sulfur, and wherein at least 60% of the heterocycles have at least one nitrogen atom,, wherein said oligomer is defined as having at least 6 of said nitrogen containing heterocycles, where at least one of said heterocycles is specific for A, G, C or T, and a complementary pair of heterocycles refers to the complementary pair of nucleotides, said oligomer comprising at least two units of 2 consecutive heterocycles forming complementary pairs with itself as a first oligomer or with another oligomer as second oligomers,  
where when said oligomer forms said complementary pairs with itself, said oligomer comprises an internal molecule forming a hairpin turn, and  
when two oligomers form said complementary pairs, said two oligomers comprise an internal aliphatic amino acid of from 2 to 6 carbon atoms, where said internal aliphatic amino acid is preferentially in juxtaposition to A and T and forms a complementary pair with itself, and  
at least one of terminal to a terminal organic cyclic group: (2) an aliphatic amino acid of from 2 to 6 carbon atoms; and (3) an alkyl chain comprising a polar group from 2 to 4 carbon atoms from the bond linking said alkyl chain to the remainder of the oligomer, said heterocycles being linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen and oxygen atoms of said dsDNA, with the proviso that hydrogen atoms away from the surface of said minor groove may be substituted with substituents of a total of not greater than 100 carbon atoms, said method comprising:  
bringing together under complex forming conditions, said oligomers with dsDNA, whereby complexes form between said first and second oligomers at any target dsDNA.
2. A method according to Claim 1, wherein said core structure is unsubstituted.
3. A method for forming a specific complex between target dsDNA in the minor groove and from 1 to 2 oligomers, where the oligomers are selected to provide binding to



said minor groove site, said oligomers comprised of N-heterocycles consisting of N-methyl pyrrole (Py) and N-methyl imidazole (Im), wherein said N-heterocycles and other members of said oligomers are selected to provide a  $K_a$  of greater than about  $10^9 \text{ M}^{-1}$ , wherein said oligomer is comprised of at least 6 said heterocycles,

5        where the order of heterocycles in relation to said target dsDNA is defined as Im/Py in juxtaposition to G/C, Py/Im in juxtaposition to C/G, and Py/Py in juxtaposition to A/T and T/A and a complementary pair of heterocycles refers to the complementary pair of nucleotides,

10        said oligomer comprising at least two units of consecutive heterocycles forming complementary pairs with itself as a first oligomer or another oligomer as second oligomers,

      where when said oligomer forms said complementary pairs with itself, said oligomer comprises an internal  $\gamma$ -aminobutyric acid, and

15        when two oligomers form said complementary pairs, said oligomers comprise an internal  $\beta$ -alanine, said internal  $\beta$ -alanine being in juxtaposition to A/T and T/A and forming a complementary pair with itself,

20        each oligomer terminating in a glycine or  $\beta$ -alanine amino acid joined to an alkyl chain of from 2 to 4 carbon atoms comprising a polar group, said heterocycles being linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen and oxygen atoms of said dsDNA, with the proviso that a second  $\gamma$ -aminobutyric acid may join the termini to define a ring of said first oligomer and hydrogen atoms away from the surface of said minor groove may be substituted with substituents of a total of not greater than 100 carbon atoms, said method comprising:

25        bringing together under complex forming conditions, said first or second oligomers with dsDNA, whereby complexes form between said first and second oligomers and any target dsDNA.

4.        A method according to Claim 3, wherein said core unit is unsubstituted.

30    5.        A method according to Claim 3, wherein said linking groups comprise amido groups.

6.        A method according to Claim 3, wherein said second oligomers have one  $\beta$ -alanine separating units of at least 2 N-heterocycles.

7. A method according to Claim 5, wherein there are not more than 2 consecutive  
Im.
8. A method according to Claim 5, wherein said second oligomers comprise at least 8  
5 N-heterocycles.
9. A method according to Claim 3, wherein at least one of said second oligomers  
comprises at least 2 unpaired N-heterocycles.
- 10 10. A method according to Claim 8, wherein each of said second oligomers comprises  
3 unpaired N-heterocycles.
11. A method according to Claim 3, wherein said first oligomer comprises at least 8  
N-heterocycles.
- 15 12. A method according to Claim 9, wherein said first oligomer comprises at least 2  
unpaired N-heterocycles.
11. A method according to Claim 3, wherein said core structure is unsubstituted.
- 20 12. A method for forming a specific complex between target dsDNA in the minor  
groove and an oligomer of N-heterocycles consisting of N-methyl pyrrole (Py) and N-  
methyl imidazole (Im), wherein said N-heterocycles are selected to provide binding in said  
minor groove with a  $K_a$  of at least  $10^9 M^{-1}$ , wherein said oligomer is comprised of at least 6  
25 said heterocycles, where the order of heterocycles in relation to said target dsDNA is  
defined as Im/Py in juxtaposition to G/C, Py/Im in juxtaposition to C/G, and Py/Py in  
juxtaposition to A/T and T/A and a complementary pair of heterocycles refers to the  
complementary pair of nucleotides,  
said oligomer comprising at least two units of 2 consecutive heterocycles forming  
30 complementary pairs with itself, said oligomer comprises an internal  $\gamma$ -aminobutyric acid  
between said two units, and a unit of 6 N-heterocycles comprises an internal  $\beta$ -alanine,  
said internal  $\beta$ -alanine being in juxtaposition to A and T and forming a complementary  
pair with itself,

said oligomer terminating in a glycine or  $\beta$ -alanine amino acid joined to an alkyl chain of from 2 to 4 carbon atoms comprising a polar group, said heterocycles being linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen and oxygen atoms of said dsDNA,

5 with the proviso that hydrogen atoms away from the surface of said minor groove may be substituted with substituents of a total of not greater than 30 carbon atoms, said method comprising:

bringing together under complex forming conditions, said oligomer with dsDNA, whereby complexes form between said oligomer and any target dsDNA.

10

13. A method according to Claim 12, wherein said polar group is a tertiary amine, with the proviso that only one terminus of an oligomer comprises said tertiary amine.

14. A method according to Claim 12, wherein said polar group is an hydroxyl group.

15

15. A method according to Claim 12, wherein said oligomer comprises at least 4 complementary pairs.

16. A method according to Claim 12, wherein two oligomers are employed, each  
20 oligomer having at least a portion of an overhang complementary to the other oligomer.

17. A method for forming a specific complex between target dsDNA in the minor groove and a pair of oligomers of N-heterocycles consisting of N-methyl pyrrole (Py) and N-methyl imidazole (Im), wherein said N-heterocycles are selected to provide binding in  
25 said minor groove with a  $K_a$  of at least  $10^9 \text{ M}^{-1}$ , wherein said oligomer is comprised of at least 6 said heterocycles, where the order of heterocycles in relation to said target dsDNA is defined as Im/Py in juxtaposition to G/C, Py/Im in juxtaposition to C/G, and Py/Py in juxtaposition to A/T and T/A and a complementary pair of heterocycles refers to the complementary pair of nucleotides,

30 said oligomers comprise an internal  $\beta$ -alanine, said internal  $\beta$ -alanine being in juxtaposition to A and T and forming a complementary pair with itself, each oligomer terminating in a glycine or  $\beta$ -alanine amino acid joined to an alkyl chain of from 2 to 4 carbon atoms comprising a polar group, said heterocycles being linked by chains of

oligomers 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen and oxygen atoms of said dsDNA,

with the proviso that hydrogen atoms away from the surface of said minor groove may be substituted with substituents of a total of not greater than 30 carbon atoms, said

5 method comprising:

bringing together under complex forming conditions, said oligomers with dsDNA, whereby complexes form between said oligomers and any target dsDNA.

18. A method according to Claim 17, wherein said oligomer comprises at least two  $\beta$ -alanines.  
10

19. A method according to Claim 17, wherein said oligomers are unsubstituted.

20. A method according to Claim 1, wherein said target dsDNA is a portion of a chromosome.  
15

21. A method according to Claim 1, wherein said target dsDNA is a portion of an episomal element.

22. A method according to Claim 1, wherein said target dsDNA is a portion of a virus.  
20

23. A method for detecting the presence of target dsDNA in a sample, employing a composition comprising from 1 to 2 oligomers at a minor groove site, where said  
25 oligomers are selected to provide binding to said target dsDNA at a  $K_a$  of at least  $10^9 M^{-1}$ ,  
said oligomers comprising as a core structure: (1) organic cyclic groups of from 5 to 6 annular members, where at least 60% of the rings are heterocycles having from 1 to 3 heteroatoms, where the heteroatoms are nitrogen, oxygen and sulfur, and wherein at least 60% of the heterocycles have at least one nitrogen atom,, wherein said oligomer is defined  
30 as having at least 6 of said nitrogen containing heterocycles, where at least one of said heterocycles is specific for A, G, C or T, and a complementary pair of heterocycles refers to the complementary pair of nucleotides, said oligomer comprising at least two units of 2 consecutive heterocycles forming complementary pairs with itself as a first oligomer or with another oligomer as second oligomers,

where when said oligomer forms said complementary pairs with itself, said oligomer comprises an internal molecule forming a hairpin turn, and

when two oligomers form said complementary pairs, said two oligomers comprise an internal aliphatic amino acid of from 2 to 6 carbon atoms, where said internal aliphatic amino acid is preferentially in juxtaposition to A and T and forms a complementary pair with itself, and

at least one of terminal to a terminal organic cyclic group: (2) an aliphatic amino acid of from 2 to 6 carbon atoms; and (3) an alkyl chain comprising a polar group from 2 to 4 carbon atoms from the bond linking said alkyl chain to the remainder of the oligomer, said heterocycles being linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen and oxygen atoms of said dsDNA, with the proviso that hydrogen atoms away from the surface of said minor groove may be substituted with substituents of a total of not greater than 100 carbon atoms, and

(4) a moiety for detecting said complex, said heterocycles being linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen atoms of said dsDNA, said method comprising:

combining said composition and said sample under complex forming conditions;

and

detecting the presence of said target dsDNA in said sample as a complex with said oligomers by means of said moiety.

24. A method of detecting target dsDNA in a sample employing from 1 to 2 oligomers, where the oligomers are selected to provide binding to a minor groove site of said target dsDNA, said oligomers comprised of N-heterocycles consisting of N-methyl pyrrole (Py) and N-methyl imidazole (Im), wherein said N-heterocycles and other members of said oligomers are selected to provide a  $K_a$  of greater than about  $10^9 M^{-1}$ , wherein said oligomer is comprised of at least 6 said heterocycles,

where the order of heterocycles in relation to said target dsDNA is defined as Im/Py in juxtaposition to G/C, Py/Im in juxtaposition to C/G, and Py/Py in juxtaposition to A/T and T/A and a complementary pair of heterocycles refers to the complementary pair of nucleotides,

said oligomer comprising at least two units of consecutive heterocycles forming complementary pairs with itself as a first oligomer or another oligomer as second oligomers,

- where when said oligomer forms said complementary pairs with itself, said oligomer comprises an internal  $\gamma$ -aminobutyric acid, and
- when two oligomers form said complementary pairs, said oligomers comprise an internal  $\beta$ -alanine, said internal  $\beta$ -alanine being in juxtaposition to A/T and T/A and
- 5 forming a complementary pair with itself,
- each oligomer terminating in a glycine or  $\beta$ -alanine amino acid joined to an alkyl chain of from 2 to 4 carbon atoms comprising a polar group, said heterocycles being linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen and oxygen atoms of said dsDNA, with the proviso that a second  $\gamma$ -
- 10 aminobutyric acid may join the termini to define a ring of said first oligomer and hydrogen atoms away from the surface of said minor groove may be substituted with substituents of a total of not greater than 100 carbon atoms,
- at least one oligomer joined to a moiety for detection of complex formation between said target dsDNA and said oligomers, said method comprising:
- 15 combining said oligomers and said sample under complex forming conditions; and detecting the presence of said target dsDNA in said sample as a complex with said oligomers by means of said moiety.
25. A method according to Claim 24, wherein said moiety is an enzyme, a fluorescer, a chemiluminescer, a solid surface, a hapten which binds to a receptor, or a radioactive isotope.
26. A method according to Claim 24, wherein said method further comprises:
- separating any complex from any other dsDNA in said sample before detecting
- 25 said complex.
27. A method according to Claim 26, wherein one of said oligomers or said dsDNA is bound to a solid surface.
- 30 28. A method according to Claim 24, wherein said moiety is biotin or digoxin.

29. A method according to Claim 24, wherein said dsDNA is chromosomal fragments.

30. A method for isolating target dsDNA from a mixture of dsDNA employing a composition comprising from 1 to 2 components, where the components are selected to provide binding to said target dsDNA of a  $K_d \leq 1 \text{ nM}$ , said components consisting of from 1 to 2 oligomers, said oligomers including: (1) organic cyclic groups of from 5 to 6 annular members, where at least 60% of the rings are heterocycles having from 1 to 3 heteroatoms, where the heteroatoms are nitrogen, oxygen and sulfur, and wherein at least 60% of the heterocycles have at least one nitrogen atom, wherein said oligomer is defined as having at least 6 of said nitrogen containing heterocycles, where at least one of said heterocycles is specific for A, G, C or T, and a complementary pair of heterocycles refers to the complementary pair of nucleotides, said oligomer comprising at least two units of 2 consecutive heterocycles forming complementary pairs with itself as a first oligomer or with another oligomer as second oligomers, where when said oligomer forms said complementary pairs with itself, said oligomer comprises an internal molecule forming a hairpin turn, and when two oligomers form said complementary pairs, said two oligomers comprise an internal aliphatic amino acid of from 2 to 6 carbon atoms, where said internal aliphatic amino acid is preferentially in juxtaposition to A and T and forms a complementary pair with itself, and at least one of terminal to a terminal organic cyclic group: (2) an aliphatic amino acid of from 2 to 6 carbon atoms; and (3) an alkyl chain comprising a polar group from 2 to 4 carbon atoms from the bond linking said alkyl chain to the remainder of the component, and (4) a moiety for separating said complex, said heterocycles being linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen atoms of said dsDNA, said method comprising:

25 combining said composition comprising said oligomers with said mixture of dsDNA under complex forming conditions; and

separating complexes which form by means of said moiety.

31. A method for isolating target dsDNA from a mixture of dsDNA employing a composition comprising from 1 to 2 oligomers at a minor groove site, where said oligomers are selected to provide binding to said target dsDNA at a  $K_a$  of at least  $10^9 \text{ M}^{-1}$ , said oligomers comprising as a core structure: (1) organic cyclic groups of from 5 to 6 annular members, where at least 60% of the rings are heterocycles having from 1 to 3 heteroatoms, where the heteroatoms are nitrogen, oxygen and sulfur, and wherein at least

60% of the heterocycles have at least one nitrogen atom,, wherein said oligomer is defined as having at least 6 of said nitrogen containing heterocycles, where at least one of said heterocycles is specific for A, G, C or T, and a complementary pair of heterocycles refers to the complementary pair of nucleotides, said oligomer comprising at least two units of 2  
5 consecutive heterocycles forming complementary pairs with itself as a first oligomer or with another oligomer as second oligomers,

where when said oligomer forms said complementary pairs with itself, said oligomer comprises an internal molecule forming a hairpin turn, and

when two oligomers form said complementary pairs, said two oligomers comprise  
10 an internal aliphatic amino acid of from 2 to 6 carbon atoms, where said internal aliphatic amino acid is preferentially in juxtaposition to A and T and forms a complementary pair with itself, and

at least one of terminal to a terminal organic cyclic group: (2) an aliphatic amino acid of from 2 to 6 carbon atoms; and (3) an alkyl chain comprising a polar group from 2  
15 to 4 carbon atoms from the bond linking said alkyl chain to the remainder of the oligomer, said heterocycles being linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen and oxygen atoms of said dsDNA, with the proviso that hydrogen atoms away from the surface of said minor groove may be substituted with substituents of a total of not greater than 100 carbon atoms, and

20 (4) a moiety for detecting said complex,  
said heterocycles being linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen atoms of said dsDNA, at least one oligomer joined to a moiety for separation of complexes between said target dsDNA and said oligomers, said method comprising:

25 combining said oligomers and said sample under complex forming conditions; and;  
separating complexes which form by means of said moiety.

32. A method of separating target dsDNA from a mixture of dsDNA employing from 1 to 2 oligomers, where the oligomers are selected to provide binding to a minor groove  
30 site of said target dsDNA, said oligomers comprised of N-heterocycles consisting of N-methyl pyrrole (Py) and N-methyl imidazole (Im), wherein said N-heterocycles and other members of said oligomers are selected to provide a  $K_a$  of greater than about  $10^9 M^{-1}$ , wherein said oligomer is comprised of at least 6 said heterocycles,



where the order of heterocycles in relation to said target dsDNA is defined as Im/Py in juxtaposition to G/C, Py/Im in juxtaposition to C/G, and Py/Py in juxtaposition to A/T and T/A and a complementary pair of heterocycles refers to the complementary pair of nucleotides,

- 5        said oligomer comprising at least two units of consecutive heterocycles forming complementary pairs with itself as a first oligomer or another oligomer as second oligomers,

where when said oligomer forms said complementary pairs with itself, said oligomer comprises an internal  $\gamma$ -aminobutyric acid, and

- 10        when two oligomers form said complementary pairs, said oligomers comprise an internal  $\beta$ -alanine, said internal  $\beta$ -alanine being in juxtaposition to A/T and T/A and forming a complementary pair with itself,

- each oligomer terminating in a glycine or  $\beta$ -alanine amino acid joined to an alkyl chain of from 2 to 4 carbon atoms comprising a polar group, said heterocycles being  
15 linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen and oxygen atoms of said dsDNA, with the proviso that a second  $\gamma$ -aminobutyric acid may join the termini to define a ring of said first oligomer and hydrogen atoms away from the surface of said minor groove may be substituted with substituents of a total of not greater than 100 carbon atoms, and

- 20        (4) a moiety for detecting said complex,  
said heterocycles being linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen atoms of said dsDNA, at least one oligomer joined to a moiety for separation of complexes between said target dsDNA and said oligomers, said method comprising:

- 25        combining said oligomers and said sample under complex forming conditions; and;  
separating complexes which form by means of said moiety.

33.    A method according to Claim 32, wherein said moiety is a hapten and said separating comprises combining said oligomers and mixture with a receptor for said hapten bound to a solid surface.

30

34.    A method according to Claim 33, wherein said solid surface is particles or a vessel wall.

35. A composition comprising from 1 to 2 oligomers where the oligomers are selected to provide binding to a minor groove site of said target dsDNA, said oligomers comprised of N-heterocycles consisting of N-methyl pyrrole (Py) and N-methyl imidazole (Im), wherein said N-heterocycles and other members of said oligomers are selected to provide
- 5 a  $K_d$  of greater than about  $10^9 \text{ M}^{-1}$ , wherein said oligomer is comprised of at least 6 said heterocycles,
- where the order of heterocycles in relation to said target dsDNA is defined as Im/Py in juxtaposition to G/C, Py/Im in juxtaposition to C/G, and Py/Py in juxtaposition to A/T and T/A and a complementary pair of heterocycles refers to the complementary
- 10 pair of nucleotides,
- said oligomer comprising at least two units of consecutive heterocycles forming complementary pairs with itself as a first oligomer or another oligomer as second oligomers,
- where when said oligomer forms said complementary pairs with itself, said
- 15 oligomer comprises an internal  $\gamma$ -aminobutyric acid, and
- when two oligomers form said complementary pairs, said oligomers comprise an internal  $\beta$ -alanine, said internal  $\beta$ -alanine being in juxtaposition to A/T and T/A and forming a complementary pair with itself,
- each oligomer terminating in a glycine or  $\beta$ -alanine amino acid joined to an alkyl
- 20 chain of from 2 to 4 carbon atoms comprising a polar group, said heterocycles being linked by chains of 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen and oxygen atoms of said dsDNA, with the proviso that a second  $\gamma$ -aminobutyric acid may join the termini to define a ring of said first oligomer and hydrogen atoms away from the surface of said minor groove may be substituted with
- 25 substituents of a total of not greater than 100 carbon atoms.

36. A composition according to Claim 35, wherein said 2 atoms of said linking groups are a carbamyl group.
- 30 37. A composition according to Claim 35, wherein said composition comprises one oligomer.
38. A composition according to Claim 37, wherein said one oligomer comprises at least 7 N-heterocycles.

39. A composition according to Claim 37, wherein said one oligomer comprises a  $\beta$ -alanine internal to six consecutive N-heterocycles and separated from said  $\gamma$ -aminobutyric acid by at least 2 heterocycles.
- 5 40. A composition according to Claim 35, wherein said composition comprises 2 oligomers, each oligomer having a sequence of at least 6 N-heterocycles and including a  $\beta$ -alanine internal to said sequence.
41. A composition according to Claim 35, wherein at least one said oligomer is  
10 substituted with a chelated metal group.
42. A composition according to Claim 35, wherein said oligomers are unsubstituted.
43. A composition according to Claim 35, wherein the total number of carbon atoms  
15 of said substituents is not greater than 30 carbon atoms.
44. A method for isolating target dsDNA from an extended dsDNA comprising said target dsDNA, said method comprising:  
bringing together said larger fragment of dsDNA and a pair of oligomers of N-  
20 heterocycles consisting of N-methyl pyrrole (Py) and N-methyl imidazole (Im), wherein said N-heterocycles are selected to provide binding in said minor groove with a  $K_a$  of at least  $10^9 M^{-1}$ , wherein said oligomer is comprised of at least 6 said heterocycles, where the order of heterocycles in relation to said target dsDNA is defined as Im/Py in juxtaposition to G/C, Py/Im in juxtaposition to C/G, and Py/Py in juxtaposition to A/T and T/A and a  
25 complementary pair of heterocycles refers to the complementary pair of nucleotides,  
said oligomers comprise an internal  $\beta$ -alanine, said internal  $\beta$ -alanine being in juxtaposition to A and T and forming a complementary pair with itself, each oligomer terminating in a glycine or  $\beta$ -alanine amino acid joined to an alkyl chain of from 2 to 4 carbon atoms comprising a polar group, said heterocycles being linked by chains of  
30 oligomers 2 atoms comprising NH groups for forming hydrogen bonds to available nitrogen and oxygen atoms of said dsDNA,  
with the proviso that hydrogen atoms away from the surface of said minor groove may be substituted with substituents of a total of not greater than 30 carbon atoms, and at at least one terminal end of an oligomer is a functionality capable of cleaving dsDNA,

whereby a complex is formed between said oligomers and said extended dsDNA;  
and  
cleaving said extended dsDNA by means of said functionality.

5 45. A method according to Claim 44, wherein said functionality capable of cleaving dsDNA is present at the same end of both oligomers and is a chelated metal.

46. A minor-groove binding polyamide reagent which binds to duplex DNA in a sequence specific manner, and has from about one to three chemical moieties appended thereto, wherein,  
10 the chemical moieties confer upon the polyamide reagent properties which render the polyamide reagent useful for a purpose selected from either purifying duplex DNA in a sequence specific manner, and detecting duplex DNA in a sequence specific manner.

47. The method of purifying duplex DNA in a sequence specific manner, by reversible  
15 immobilization, which employs the polyamide reagent of Claim 46.

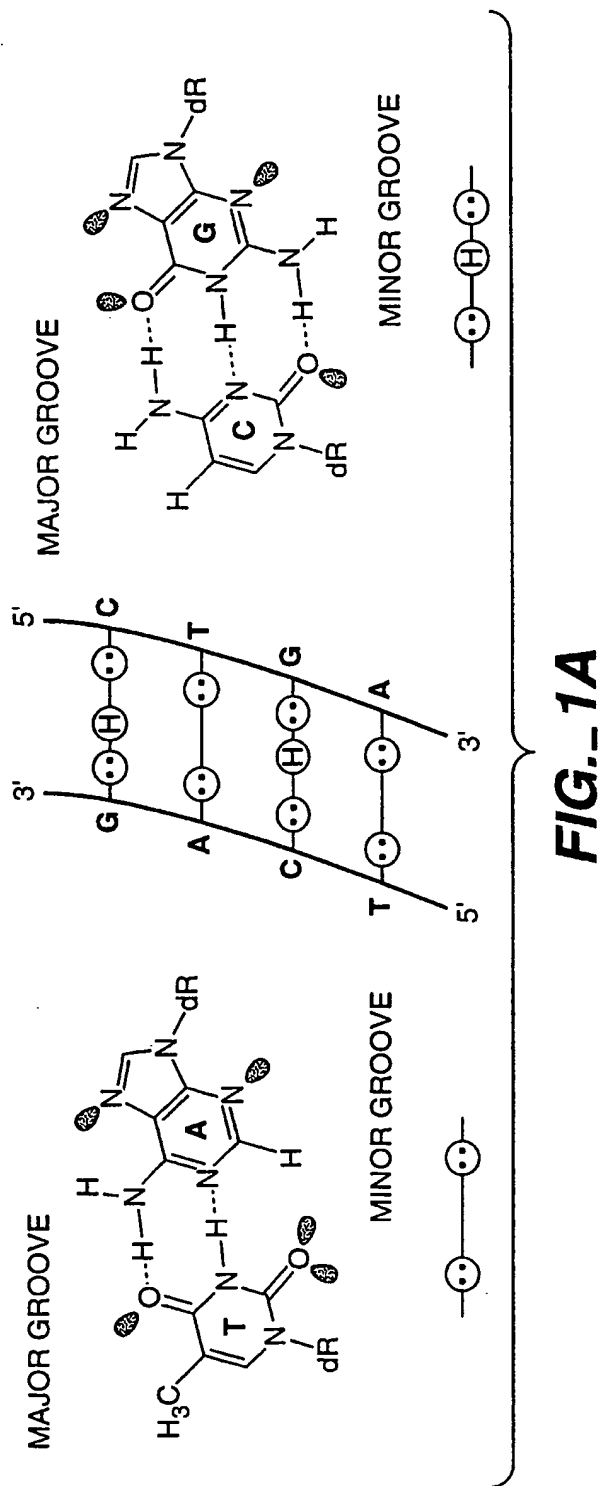
48. The method of Claim 47, wherein, the chemical moieties appended to the polyamide reagent are selected from a group comprised of either arylboronic acids, biotins, polyhistidines comprised of from about 2 to 8 amino acids, haptens to which an antibody binds, and solid phase  
20 supports.

49. The method of detecting duplex DNA in a sequence specific manner which employs the polyamide reagent of Claim 46.

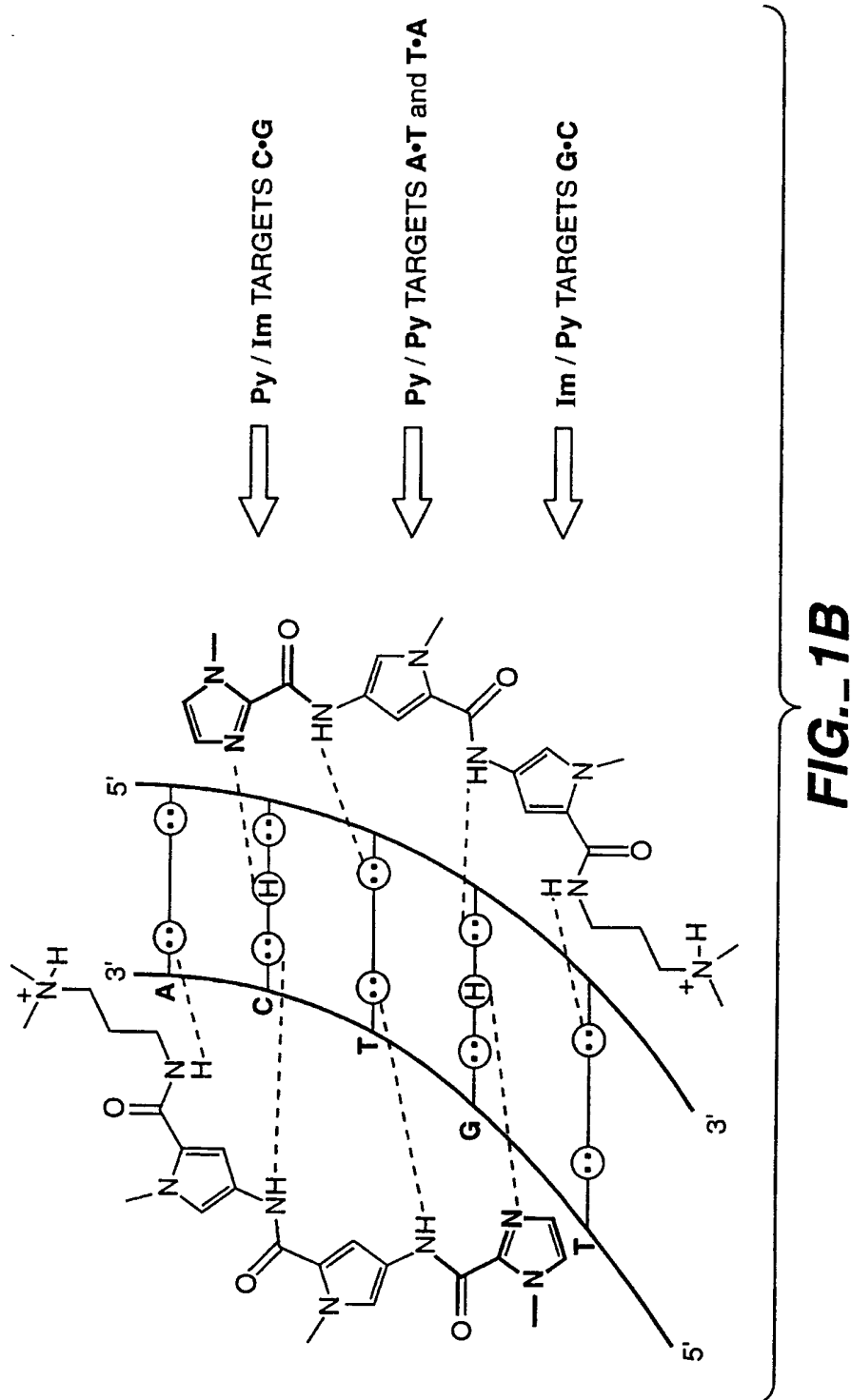
25 50. The method of Claim 49, wherein, the chemical moieties appended to the polyamide reagent are selected from a group comprised of either chromophores, fluorophores, metal ion chelators, enzymes, or other moieties detectable by visual, spectroscopic or electronic means.

51. The method of Claim 50, wherein, the chemical moieties appended to the polyamide  
30 reagent are comprised of two fluorophores which act as an energy donor and energy acceptor pair.

1 / 6



2 / 6



3/6

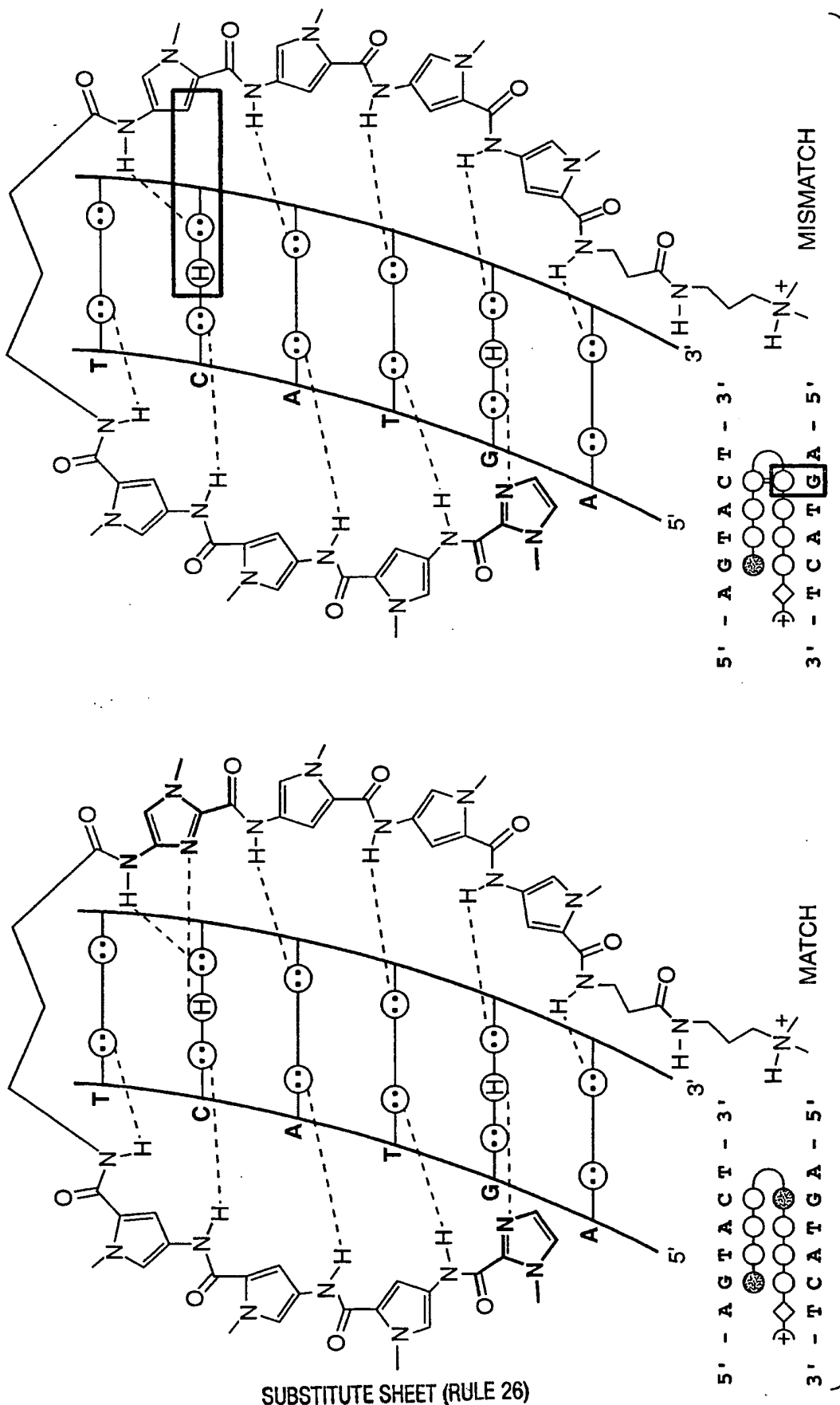


FIG. 2A

SUBSTITUTE SHEET (RULE 26)

4 / 6

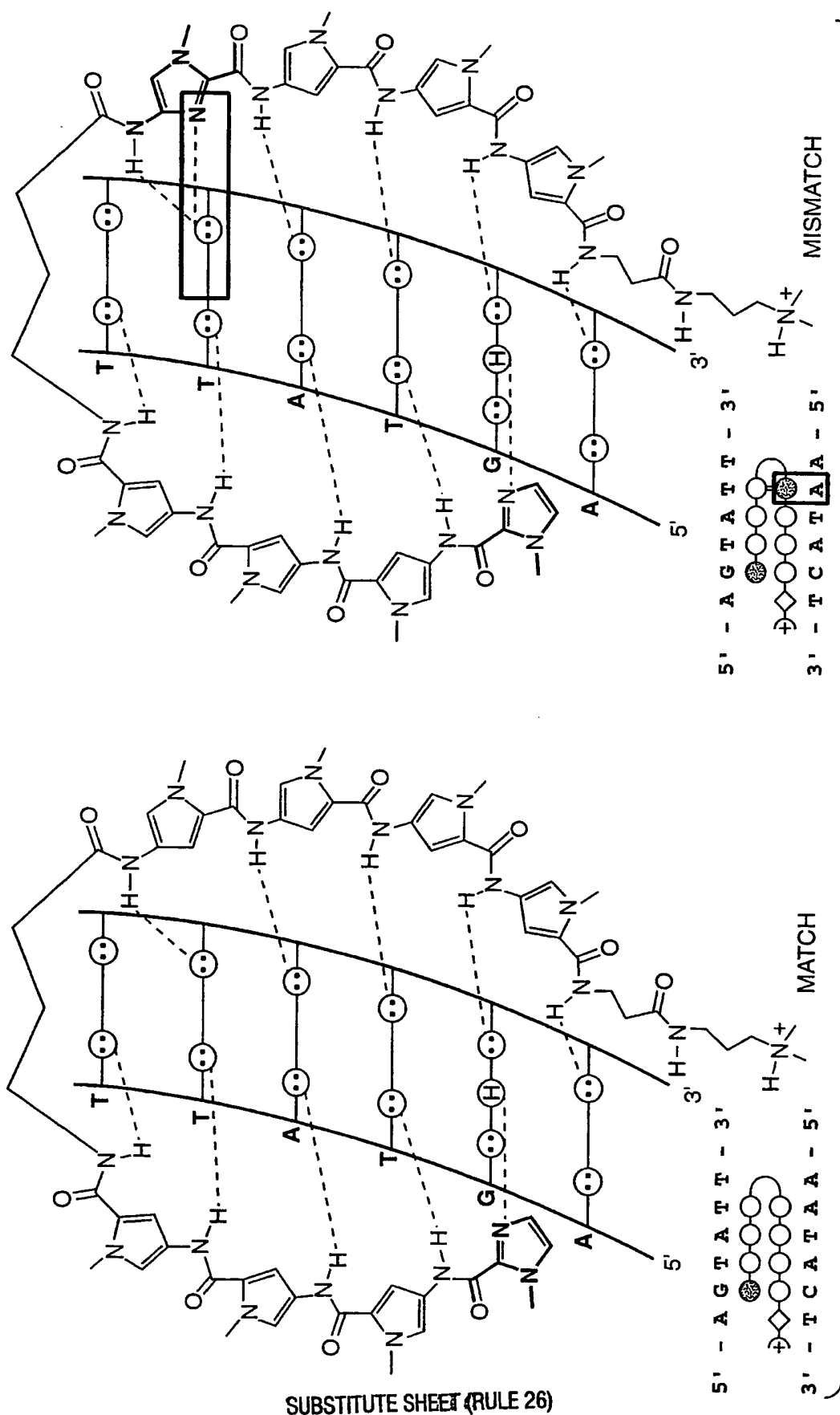
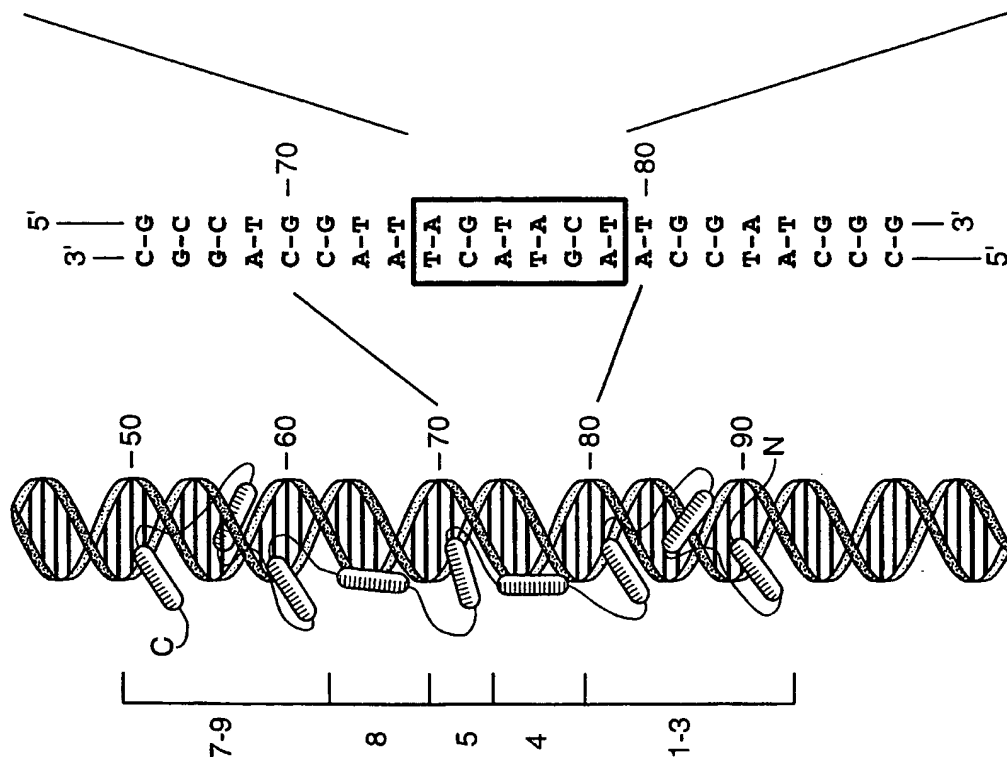
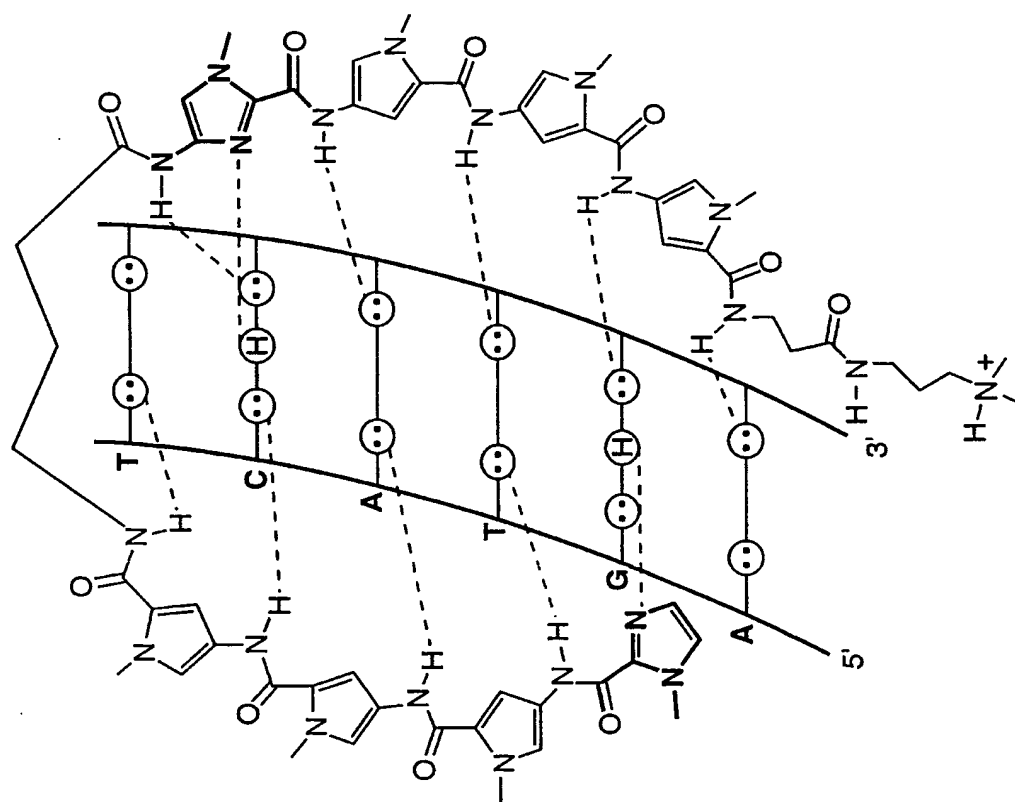


FIG. 2B



5 / 6



SUBSTITUTE SHEET (RULE 26)

FIG. 3A

6/6

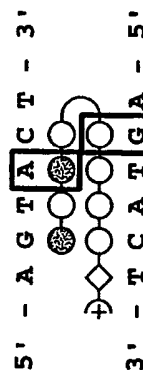
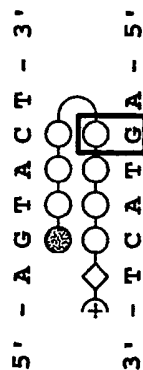
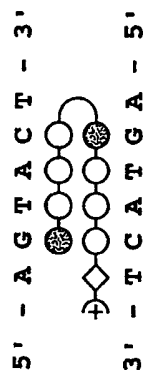


FIG. 3C

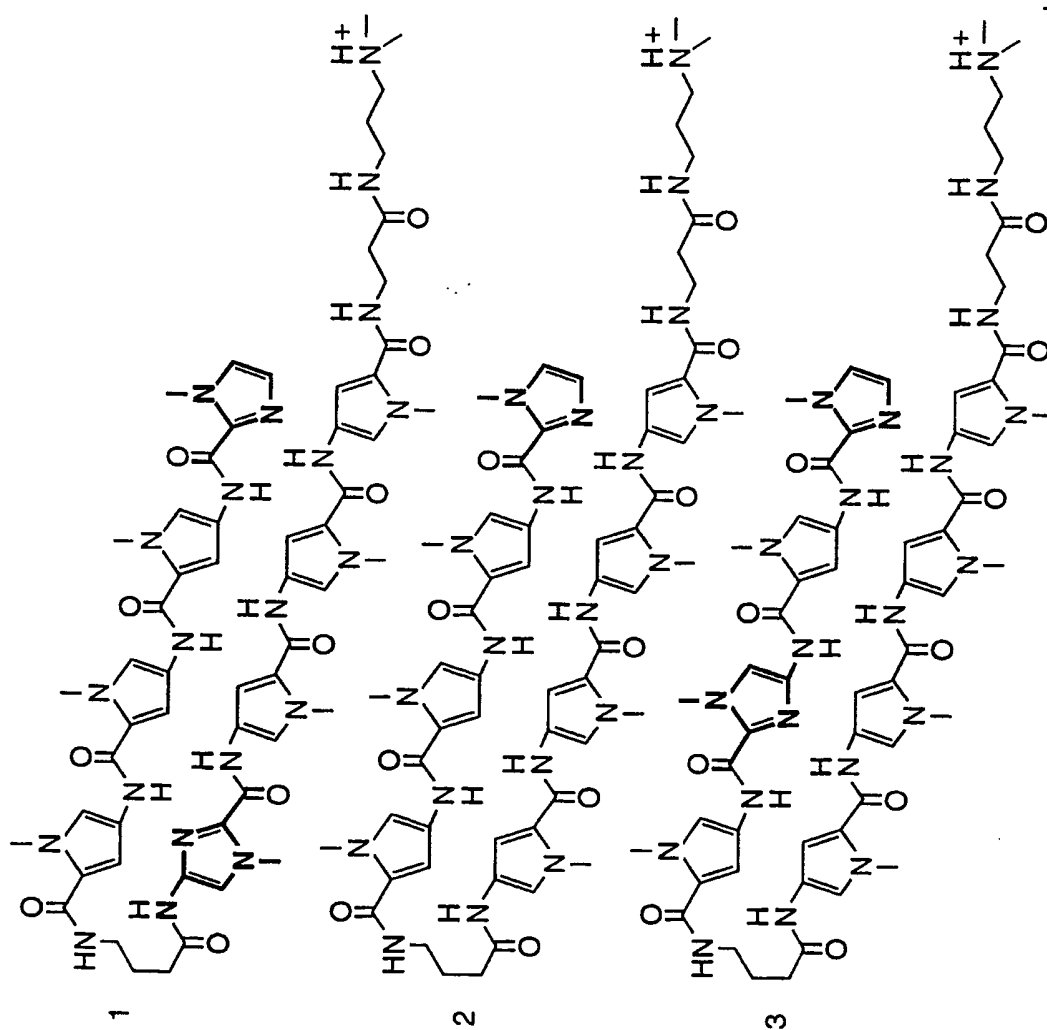


FIG. 3B

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/12722**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(6) :C12Q 1/68, 1/70; C12P 19/34; C07H 21/04, 21/02; A61K 41/00

US CL :435/6, 5, 91.1; 536/23.1, 24.3, 24.32, 24.33, 24.5; 514/44

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 5, 91.1; 536/23.1, 24.3, 24.32, 24.33, 24.5; 514/44

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 4,795,700 A (DERVAN et al) 03 January 1989, see abstract.	41,44
A	US 5,616,606 A (LOWN ET AL.) 01 April 1997, see abstract.	1-51
A	US 4,912,199 A (LOWN et al) 27 March 1990, see abstract.	1-51
A	US 5,459,255 A (COOK et al) 17 October 1995, see entire document.	1-51

☐ Further documents are listed in the continuation of Box C.
 ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

03 OCTOBER 1997

Date of mailing of the international search report

31 OCT 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231Authorized officer  
DIANNE REES

Facsimile No. (703) 305-3230

Telephone No. (703) 308-0196

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/12722

### B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, BIOSIS, BIOTECHABS, BIOTECHDS, CABA, CAPLUS, CJELSEVIER, CANCERLIT, AIDSLINE, DISSABS, EMBASE, MEDLINE, EUROPATFULL, EUROPEX, SCISEARCH, DRUGU, TOXLINE, TOXLIT, JAPIO, WPIDS, USPATFULL

search terms: minor groove, heterocycles, heterocyclic,, nitrogen, oxygen, sulfur, hairpin,liip, aliphatic amino acid, biding, affinity, hybridization, hybridisation, N-methyl pyrrole, Py, N-methyl imidazole, Im, aminobutyric acid, polar groups,amido groups,,alanine, glycine, , chromosome, episomal, extrachromosomal, virus, , rings, , isolating, separating, solid supports,carbamyl, linking groups, chelator, metal, . cleaving, digestig, cleavage, digestion,, haptens, antibody,chromophores, fluorophores, metal ion chelators, enzymes, energy donar, acceptor, quencher, quenched